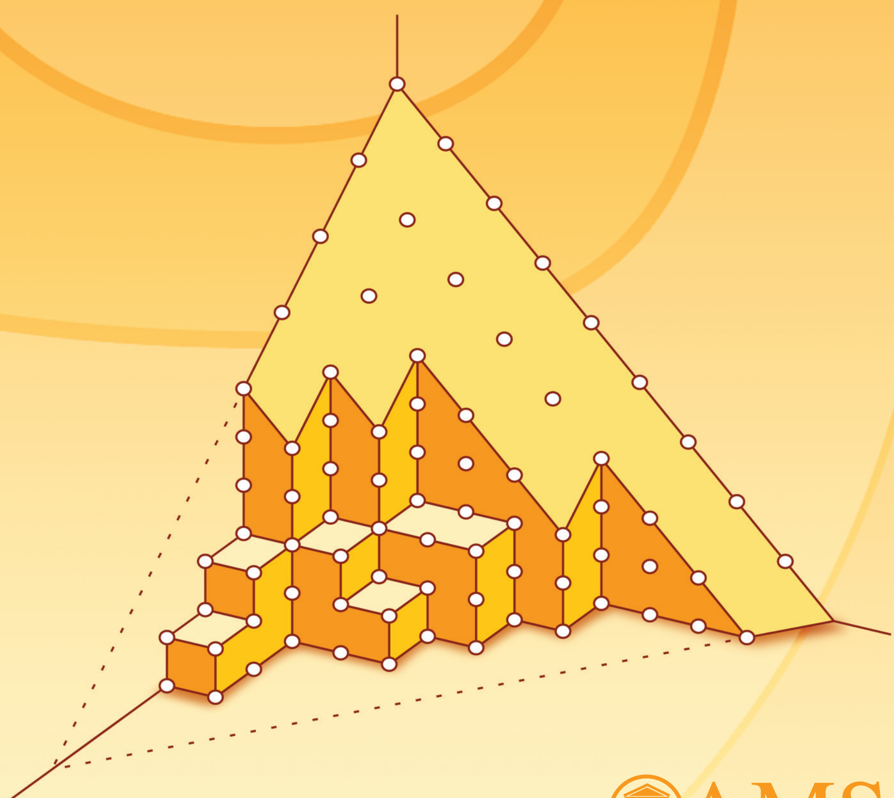
STUDENT MATHEMATICAL LIBRARY
Volume 75

Mathematics++

Selected Topics Beyond the Basic Courses

Ida Kantor
Jiří Matoušek
Robert Šámal



AMS
AMERICAN MATHEMATICAL SOCIETY

VISIT…

LANZAROTE Caliente.com

Mathematics++

Selected Topics Beyond the Basic Courses

STUDENT MATHEMATICAL LIBRARY
Volume 75

Mathematics++

Selected Topics Beyond the Basic Courses

Ida Kantor
Jiří Matoušek
Robert Šámal

AMS
American Mathematical Society
Providence, Rhode Island

Editorial Board

Satyan L. Devadoss — John Stillwell (Chair)
Erica Flapan — Serge Tabachnikov

2010 *Mathematics Subject Classification*. Primary 14-01; 20Cxx, 28-01, 43-01, 52Axx, 54-01, 55-01.

For additional information and updates on this book, visit
www.ams.org/bookpages/stml-75

Library of Congress Cataloging-in-Publication Data

Kantor, Ida, 1981-
Mathematics++ : selected topics beyond the basic courses / Ida Kantor, Jiří Matoušek, Robert Šámal.
pages cm. – (Student mathematical library ; volume 75)
Includes bibliographical references and index.
ISBN 978-1-4704-2261-5 (alk. paper)
1. Mathematics–Study and teaching (Graduate) 2. Computer science–Mathematics. I. Matoušek, Jiří, 1963–2015 II. Šámal, Robert, 1977- III. Title.

QA11.2.K36 2015
510–dc23 2015016136

Copying and reprinting. Individual readers of this publication, and nonprofit libraries acting for them, are permitted to make fair use of the material, such as to copy select pages for use in teaching or research. Permission is granted to quote brief passages from this publication in reviews, provided the customary acknowledgment of the source is given.

Republication, systematic copying, or multiple reproduction of any material in this publication is permitted only under license from the American Mathematical Society. Permissions to reuse portions of AMS publication content are handled by Copyright Clearance Center's RightsLink® service. For more information, please visit: `http://www.ams.org/rightslink`.

Send requests for translation rights and licensed reprints to `reprint-permission@ams.org`.

Excluded from these provisions is material for which the author holds copyright. In such cases, requests for permission to reuse or reprint material should be addressed directly to the author(s). Copyright ownership is indicated on the copyright page, or on the lower right-hand corner of the first page of each article within proceedings volumes.

©2015 American Mathematical Society. All rights reserved.
Printed in the United States of America.

∞ The paper used in this book is acid-free and falls within the guidelines established to ensure permanence and durability.
Visit the AMS home page at `http://www.ams.org/`

10 9 8 7 6 5 4 3 2 1 20 19 18 17 16 15

Contents

Preface ix

Chapter 1. Measure and Integral 1

§1. Measure 6

§2. The Lebesgue Integral 22

§3. Foundations of Probability Theory 31

§4. Literature 36

Bibliography 37

Chapter 2. High-Dimensional Geometry and Measure Concentration 39

§1. Peculiarities of Large Dimensions 41

§2. The Brunn–Minkowski Inequality and Euclidean Isoperimetry 44

§3. The Standard Normal Distribution and the Gaussian Measure 53

§4. Measure Concentration 61

§5. Literature 81

Bibliography 82

Chapter 3. Fourier Analysis 85

§1. Characters 87

§2. The Fourier Transform 94

§3. Two Unexpected Applications 99
§4. Convolution 106
§5. Poisson Summation Formula 109
§6. Influence of Variables 113
§7. Infinite Groups 124
§8. Literature 137
Bibliography 138

Chapter 4. Representations of Finite Groups 141
§1. Basic Definitions and Examples 142
§2. Decompositions into Irreducible Representations 145
§3. Irreducible Decompositions, Characters, Orthogonality 150
§4. Irreducible Representations of the Symmetric Group 160
§5. An Application in Communication Complexity 163
§6. More Applications and Literature 168
Bibliography 169

Chapter 5. Polynomials 173
§1. Rings, Fields, and Polynomials 173
§2. The Schwartz–Zippel Theorem 175
§3. Polynomial Identity Testing 176
§4. Interpolation, Joints, and Contagious Vanishing 180
§5. Varieties, Ideals, and the Hilbert Basis Theorem 185
§6. The Nullstellensatz 188
§7. Bézout's Inequality in the Plane 195
§8. More Properties of Varieties 200
§9. Bézout's Inequality in Higher Dimensions 219
§10. Bounding the Number of Connected Components 226
§11. Literature 232
Bibliography 232

Chapter 6. Topology 235
§1. Topological Spaces and Continuous Maps 236
§2. Bits of General Topology 240
§3. Compactness 247
§4. Homotopy and Homotopy Equivalence 253
§5. The Borsuk–Ulam Theorem 257
§6. Operations on Topological Spaces 262
§7. Simplicial Complexes and Relatives 271
§8. Non-embeddability 283
§9. Homotopy Groups 289
§10. Homology of Simplicial Complexes 301
§11. Simplicial Approximation 309
§12. Homology Does Not Depend on Triangulation 314
§13. A Quick Harvest and Two More Theorems 318
§14. Manifolds 321
§15. Literature 330
Bibliography 330

Index 333

Preface

This book introduces six selected areas of mostly 20th century mathematics. We assume that the reader has gone through the usual undergraduate courses and is used to rigorous presentation with proofs.

Mathematics is beautiful, and useful all over, but extensive. Even in computer science, one of the most mathematical fields besides mathematics itself, university curricula mostly teach only mathematics developed prior to the 20th century (with the exception of areas more directly related to computing, such as logic or discrete mathematics). This is not because of lack of modernity, but because building proper foundations takes a lot of time and there is hardly room for anything else, even when the mathematical courses occupy the maximum politically acceptable part of the curriculum.

This observation was the starting point of a project resulting in this book. Contemporary research in computer science (but in other fields as well) uses numerous mathematical tools not covered in the basic courses. We had the experience of struggling with papers containing mathematical terminology unknown to us, and we saw a number of other people having similar problems.

We decided to teach a course, mainly for Ph.D. students of theoretical computer science, introducing various mathematical areas in a concise and accessible way. With expected periodicity of three semesters, the course covered one to three areas per semester.

This book is a significantly extended version of our lecture notes. In six chapters, it deals with measure theory, high-dimensional geometry, Fourier analysis, representations of groups, multivariate polynomials (or, in fancier terms, rudimentary algebraic geometry), and topology. The chapters are independent of one another and can be studied in any order.

In the selection of the areas, the reader has to rely on our experience and opinions (although we have also asked colleagues and students what they would consider most helpful for their work). This, of course, is subjective and someone else would perhaps recommend different areas, but we believe that it is at least a reasonable suggestion—and better than no suggestion.[1]

For each of the areas, we aim at presenting basic notions, basic examples, and basic results. Exercises form an integral part of the presentation, since we believe that the only way of really grasping sophisticated notions is to actively work with them. Results too advanced to be developed rigorously in a limited space but too nice to be omitted are described, sometimes slightly informally, in encyclopedia-like passages.

Since our goal was an introductory textbook, we try to keep citations to a (reasonable) minimum. We also do not always recall notions that are usually treated on undergraduate level, since we do not want to clutter the text, and since nowadays it is easy to look up definitions in trustworthy Internet resources—which we encourage the readers to do when in doubt.

There is only so much information one can fit in a single chapter; moreover, we do not try to write as compactly as possible, preferring accessibility. Even this limited amount of knowledge can often be sufficient—as a rule, among results and notions in a given field, the simpler ones have a greater chance to be applied in other fields than the advanced and fancy ones.

[1] There is a small Internet company, whose main component is apparently an experienced man scanning and reading lots of articles from world's leading periodicals every day. For a modest yearly fee, the company then provides access to his recommendations of most interesting and important articles. We believe that, when practically any opinion can be "confirmed" by some Internet pages, such trust-based services will become more and more valuable.

Once it becomes clear that one needs to know more, one can go the "standard" way, i.e., to take a full-fledged course for mathematicians or to study a textbook. This represents a considerably greater time investment, though, and very few people have the time and energy to pursue this approach for more than two or three areas.

A possible suggestion of what to add to our list of topics is probability theory. However, we believe that probability is so crucial and widely used that it is definitely worth taking a course of one, or better, two semesters.

Prerequisites. As was mentioned above, we assume that the reader has gone through the basic mathematical courses with rigorous development, including proofs. Sometimes we use bits of mathematical analysis, discrete mathematics, and very basic probability, but by far the most important background we expect is linear algebra, including vector spaces and linear maps—indeed, in current research, one encounters linear algebra at every corner.

Readership. Our course was targeted at Ph.D. students in theoretical computer science and in discrete mathematics; the applications we present are drawn from these areas.

However, the book can be useful for a much wider audience, such as mathematicians specialized in other areas, mathematics students deciding what specialization to pursue and/or preparing for graduate school, or experts in engineering or other fields using advanced mathematics. Readers not familiar with the topics of our examples are invited to look at them, but if they find some of them incomprehensible, they are free to skip such parts.

Conventions. The main notions are set in **boldface**, less important or only tangentially mentioned ones in *italics*. Exercises are interspersed through the text, and each ends with the ✠ symbol. In the index, you will find "Fubini's theorem" both under T and under F. Mathematical symbols beginning with fixed letters, such as B^n or $\mathrm{GL}(n, \mathbb{R})$ are indexed alphabetically. Notation including special symbols, such as $X \cong Y$ or $\Bbbk[X]$ (note that we may also have $A \cong Z$ or $\mathbb{R}[Y]$, so no letters are fixed), and Greek letters are listed at the beginning of the index.

Acknowledgments. Since the chapters are self-contained, we have usually asked people to read through just one, and accordingly, most acknowledgments are postponed to the ends of the chapters.

Globally we would like to thank Tomáš Toufar for proofreading the almost-finished book.

Despite the efforts, no book is ever perfect. We would appreciate to learn about mistakes and suggestions of how to improve the exposition.

It was a great pleasure to work with people from the AMS Publishing. Especially Ina Mette was incredibly helpful, and Barbara Beeton was a model of an efficient TeX expert.

We dedicate this book to our families.

Chapter 1

Measure and Integral

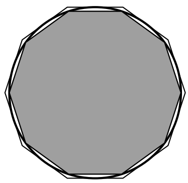
Measure theory stems from the ancient question of measuring areas. It provides definitions of lengths, areas, and volumes under very general circumstances. It is needed for building rigorous foundations of probability theory, and also for modern theory of integration (well, more modern than the 19th century).

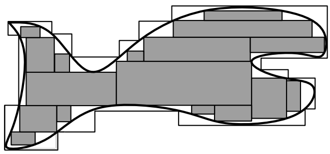
Measure theory is also a basic tool in many discrete and information-theoretic applications. Even if its fine points are seldom encountered in such contexts, it provides solid foundations and indispensable vocabulary.

Defining areas and volumes. What is the area of a possibly complicated geometric figure? This question is definitely not easy.

The area of a rectangle is the product of its sidelengths—that is taken as a basic principle (axiom). It is known that every planar polygon can be dissected into finitely many parts that can be rearranged to make a rectangle, and this can be used to define areas of polygons. However, for the circular disk such a dissection is obviously impossible, and Dehn's famous solution of Hilbert's third problem shows that even some polyhedra, such as the regular tetrahedron, cannot be dissected into finitely many parts that reassemble to a rectangular box.

The area of planar figures such as the disk, and the volume of a tetrahedron and other bodies, can be defined and computed by a limit process, the so-called *method of exhaustion*, which was already known to ancient Greek geometers (Antiphon, Eudoxus). For the disk, for example, one can inscribe and circumscribe regular polygons with larger and larger number of sides, as in the left picture:

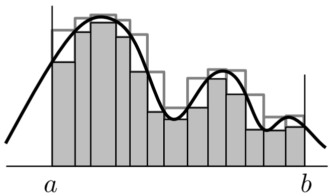




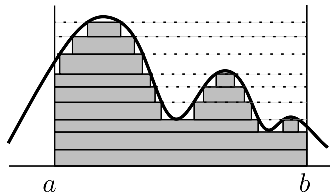
This approach, in more modern form, leads to the notion of *Jordan content* (or Jordan–Peano volume) of a set A, where one approximates A from inside and from outside by a union of finitely many non-overlapping rectangles, as is indicated in the right picture.

The Jordan content of A is defined only if the supremum of the areas of the inner approximations equals the infimum of the outer ones. Unfortunately, this leaves out many natural and important sets of everyday mathematics: for example, the set of all points in the unit square with rational coordinates has undefined Jordan content, because all inner approximations yield 0, while each outer one yields at least 1.

Integration. With the discovery of calculus, many areas and volumes could be computed by integration. Introductory courses almost always present Riemann's definition of integral, where the area under the graph of a nonnegative function f in an interval $[a, b]$ is approximated from the inside and from the outside by finer and finer vertical rectangles, as in the left picture.



Riemann



Lebesgue

This is similar to the Jordan content.

Let us note that integrals that can be computed exactly are mostly evaluated symbolically, using antiderivatives, but Riemann's definition is still needed as a formal underpinning of this method.

Here we will introduce a more sophisticated definition of areas and volumes, the *Lebesgue measure* and its abstraction to general *measure spaces*, and a better definition of integral, also due to Lebesgue.

Compared to the Riemann integral, the main conceptual change in Lebesgue's definition is that, instead of cutting vertically, he cuts *horizontally*, or in more abstract terms, instead of subdividing the domain of the integrated function, he subdivides the *range*. A nice metaphor, due to Lebesgue himself, is about counting coins in one's pocket: in the Riemann way, one takes the coins one by one and adds their values; the Lebesgue way is to sort the coins by their denominations and count the number for each denomination.

In the case of integrating a nonnegative function f as in the picture, Lebesgue approximated $\int_a^b f(x)\,\mathrm{d}x$ by the sum of the areas of finitely many horizontal slices. The contribution of the slice between the horizontal lines $y = y_{i-1}$ and $y = y_i$ should be $y_i - y_{i-1}$ times the length of the upper boundary of the slice, i.e., of the set $\{x \in [a,b] : f(x) \geq y_i\}$.

This 1-dimensional set need not be a single interval, or even a finite union of intervals—for a wild function f it can be quite complicated. This is why Lebesgue's definition of integral also needs a satisfactory definition of "length" for very complicated sets in $\mathbb{R}$. (For the Riemann integral this issue does not arise—there we deal only with areas of rectangles.) Similarly, to integrate very general functions on $\mathbb{R}^2$, we need to measure areas of very general planar sets, and so on for higher dimensions.

Fortunately, a suitable notion was already available at the time of Lebesgue's work—although nowadays we call it the Lebesgue measure, it was mostly worked out by Borel and his predecessors. While for the Jordan content we approximate by finite unions of rectangles, for the Lebesgue measure we take *countably many* rectangles (we also approximate only from outside, but that difference is less significant). This seemingly minor change allows one to use the definition for much more general sets, and the resulting notion is better behaved than the Jordan content.

Riemann vs. Lebesgue. For integrating a bounded function f defined on a closed interval (of finite length), Lebesgue's definition extends Riemann's in the sense that whenever the Riemann integral of f exists, the Lebesgue one exists as well and has the same value.[1] Similar theorem holds in $\mathbb{R}^n$.

The most famous example of a function having the Lebesgue integral but not the Riemann one is given by

$$f(x) = \begin{cases} 1 & \text{for } x \text{ rational,} \\ 0 & \text{for } x \text{ irrational.} \end{cases}$$

It is clear that $\int_0^1 f(x)\,\mathrm{d}x$ in the Riemann sense does not exist, since any inner rectangle has height 0 but any outer rectangle height 1. Later on, we will see that this integral is defined in the Lebesgue sense and equals 0.

A skeptic may regard this example as an irrelevant curiosity. It should be honestly admitted that for most functions on $\mathbb{R}^n$ one may encounter "in practice," in particular for all continuous functions, the Riemann integral works. Then, why bother with Lebesgue's more complicated definition? After all, practically all introductory courses prefer the Riemann integral because of its simplicity.

There are (at least) two main reasons for using the Lebesgue integral.

[1] The assumption that the interval is closed and bounded is not too restrictive, since a Riemann integral is naturally defined only on such intervals. These integrals are sometimes called *proper* to distinguish them from *improper* Riemann integrals, which are defined on open or unbounded intervals, as limits of proper integrals. Once we define Lebesgue integral, we will see that it may happen that an improper Riemann integral exists, while Lebesgue does not—for example, $\int_0^\infty \frac{\sin x}{x}\,\mathrm{d}x$.

- (Better properties) Even when working with specific functions, we often need to apply general theorems, and theorems about the Riemann integral are weaker and have much more complicated assumptions.

  A specific reason leading to Lebesgue's work was that the Riemann integral behaves badly under taking limits, one of the basic devices in dealing with functions. That is, if $f_1, f_2, \ldots$ is a sequence of functions and f is their pointwise limit, i.e., $f(x) = \lim_{n\to\infty} f_n(x)$ for every x, then ideally one would like to have

$$\lim_{n\to\infty} \int_a^b f_n(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x.$$

  For the Riemann integral this may fail even in seemingly innocent cases, e.g., all the f_n continuous and uniformly bounded on $[a,b]$. On the other hand, for the Lebesgue integral it holds in considerable generality, although some mild assumptions are needed.

- (Integration on general spaces) The Riemann integral is easy to define on $\mathbb{R}$, or on $\mathbb{R}^n$, but for functions on curved surfaces, for example, it becomes problematic. In contrast, the Lebesgue integral is defined on any *measure space*, i.e., whenever we specify a way of measuring "volumes" in an abstract sense.

  This abstract theory of integration was developed only many years after Lebesgue's work, and since the 1930s it serves as a foundation of modern probability theory.

Exercise 0.1. Find a sequence of Riemann integrable functions (in this exercise they need not be continuous) $f_1, f_2, \ldots : [0,1] \to [0,1]$ such that the pointwise limit $f = \lim_{n\to\infty} f_n$ exists but is not Riemann integrable. ✠

Let us remark that there is an alternative path to building the Lebesgue integral and measure: the *Daniell integral*. This integral is defined axiomatically, and the measure is then obtained as a by-product. Daniell's way has some advantages over the Lebesgue approach, e.g., greater generality. However, since for applications in

computer science and combinatorics, understanding measure seems more important than fine points of integration, and since the advantages of the Daniell integral really show only in fairly advanced contexts, we prefer the more traditional way, constructing the measure first.

1. Measure

As was indicated above, we want to define measure before defining integral. We begin with measuring subsets of the real line $\mathbb{R}$. This case may look much simpler at first sight than measuring areas in the plane, but actually, the main issues are already present in this setting.

Our goal is thus to define a sensible notion of "length" for all "reasonable" subsets $A \subseteq \mathbb{R}$. Ideally, we would of course prefer to define length for *every* subset of $\mathbb{R}$, but it turns that this is not possible to do in a sensible way, for reasons discussed later.

1.1. The Outer Lebesgue Measure. First we define the *outer Lebesgue measure*: this is a function λ^* that assigns a nonnegative real number, or ∞, to *every* set $A \subseteq \mathbb{R}$. The definition uses countable covers of A by open intervals:

Definition 1.1. *The* **outer Lebesgue measure** *of a set $A \subseteq \mathbb{R}$ is defined as*

$$\lambda^*(A) = \inf \left\{ \sum_{i=1}^{\infty} \ell(I_i) : I_1, I_2, \ldots \text{ open intervals, } A \subseteq \bigcup_{i=1}^{\infty} I_i \right\},$$

where $\ell(I)$ denotes the length of the interval I; for $I = (a, b)$, $a < b$, we have $\ell(I) = b - a$.

The values attained by λ^* include $+\infty$, and so it is convenient to say that the range of λ^* are the *extended real numbers*, denoted by $\overline{\mathbb{R}}$. As a set, $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$, and the arithmetic operations are defined as expected; e.g., $17 + \infty = \infty$. Some expressions, such as $\infty - \infty$, remain undefined.

Some good properties of λ^*. First let us check that λ^* measures the simplest sets, *intervals*, in the intended way. This is not that easy to prove, actually: one needs to use the fact that every bounded subset of $\mathbb{R}$ has a supremum, or some equivalent property, e.g., compactness of closed intervals $[a, b]$ (see Chapter 6, Section 3).

A terminological remark is in order here—an interval can be closed, half-closed, or open. It can also be finite (i.e., both endpoints are real numbers) or infinite (at least one of the endpoints is $+\infty$ or $-\infty$). We will repeatedly use that fact that λ^* is *monotone*, i.e., $A \subseteq B$ implies $\lambda^*(A) \leq \lambda^*(B)$. This follows easily from the definition.

Lemma 1.2. *If I is an interval, then $\lambda^*(I) = \ell(I)$.*

Proof. We will only prove the lemma for closed finite intervals $I = [a, b]$ and leave the rest to the reader as an exercise.

Choose a positive real ε. Since $I \subseteq (a - \varepsilon, b + \varepsilon)$, we have $\lambda^*([a, b]) \leq b - a + 2\varepsilon$. The value of ε can be chosen arbitrarily small, and so $\lambda^*([a, b]) \leq b - a$.

To prove the other inequality, we need to show that whenever $\{I_1, I_2, \ldots\}$ is a collection of open intervals covering the interval $[a, b]$, then $\sum \ell(I_i) \geq b - a$. The famous *Heine–Borel theorem* implies that there exists a *finite* subcollection of the intervals that also covers $[a, b]$ (this is the place where one needs to use some deep properties of $\mathbb{R}$). By deleting the unused intervals, the sum $\sum \ell(I_i)$ can only decrease. We can therefore suppose without loss of generality that $\{I_1, I_2, \ldots\}$ is finite.

One of the intervals I_i contains the point a; let us call it (a_1, b_1). If $b_1 < b$, then some interval, say (a_2, b_2), contains b_1. We continue in a similar fashion until we arrive at an interval (a_k, b_k) containing b. We have

$$\sum \ell(I_i) \geq (b_1 - a_1) + (b_2 - b_1) + \cdots + (b_k - b_{k-1}) = b_k - a_1 > b - a.$$

□

We strongly recommend doing the next exercise before reading further. It shows, in particular, that the set of all rational numbers

in $[0,1]$ has outer measure 0. This illustrates that the result proved above, $\lambda^*(I) = \ell(I)$, has to rely on some property that differentiates the reals from the rationals. To appreciate this issue, it may be good to remember that the discovery of irrational numbers in ancient Greece was a great surprise and one of the peak intellectual achievements of that time.

Exercise 1.3. Let $A \subset [0,1]$ be countable. Prove that $\lambda^*(A) = 0$. ✠

Next, we will check that if a set A is covered by finitely many, or even countably many, subsets, then the outer measure of A cannot be larger than the sum of the outer measures of the subsets. This property is called **countable subadditivity** of λ^*, and it immediately solves the previous exercise.

Lemma 1.4. *If $A_1, A_2, \dots$ is a countable system of subsets of $\mathbb{R}$, then*

$$\lambda^*\Big(\bigcup_{i=1}^{\infty} A_i\Big) \le \sum_{i=1}^{\infty} \lambda^*(A_i).$$

Proof. The union of sufficiently small covers of the sets A_i is a sufficiently small cover of the set $\bigcup A_i$. More precisely, if $\{I_j^{(i)}\}_{j=1}^{\infty}$ is a cover of A_i such that $\sum_j \ell(I_j^{(i)}) \le \lambda^*(A_i) + \frac{\varepsilon}{2^i}$, $i = 1, 2, \dots$, then

$$\lambda^*\Big(\bigcup A_i\Big) \le \sum_{i,j} \ell(I_j^{(i)}) \le \sum_i \lambda^*(A_i) + \varepsilon.$$

This is true for every $\varepsilon > 0$. □

The failure of additivity. Now one may ask, why not take the outer measure λ^* as the desired notion of "length"? The reason is that one would like to have length at least **finitely additive**, meaning that the length of the union of finitely many *disjoint* sets equals sum of their lengths. But for the outer measure λ^* this may fail!

Here is an example showing the failure of a stronger property, **countable additivity**.

Example 1.5. *There exist countably many sets $A_1, A_2, \dots \subset \mathbb{R}$, mutually disjoint, such that*

$$\sum_{i=1}^{\infty} \lambda^*(A_i) \ne \lambda^*\Big(\bigcup_{i=1}^{\infty} A_i\Big).$$

Proof. Call two numbers $a, b \in [0,1]$ *equivalent* if $a - b$ is rational. Let $V \subset [0,1]$, a *Vitali set*, be a set containing precisely one number from each equivalence class.

Let us enumerate all the rational numbers in $[-1,1]$ in a sequence $q_1, q_2, q_3, \ldots$, and let $A_i := V + q_i$ (the translation of V by q_i).

The A_i are clearly disjoint and contained in $[-1,2]$, and some thought reveals that they together cover $[0,1]$. Hence $\lambda^*([0,1]) = 1 \le \lambda^*(\bigcup_{i=1}^{\infty} A_i) \le \lambda^*([-1,2]) = 3$.

Now λ^* is, by definition, translation-invariant, and so $\lambda^*(A_i) = \lambda^*(V)$ for every i. So if $\lambda^*(V) = 0$, then $\sum_{i=1}^{\infty} \lambda^*(A_i) = 0$, and if $\lambda^*(V) > 0$, then $\sum_{i=1}^{\infty} \lambda^*(A_i) = \infty$. Even without knowing which of these possibilities actually holds, we can say for sure that $\sum_{i=1}^{\infty} \lambda^*(A_i)$ cannot be between 1 and 3. □

From the failure of countable additivity one can also derive the failure of finite additivity.

Exercise 1.6. Suppose (falsely!) that λ^* is finitely additive. Show that it then must be countably additive as well. ✠

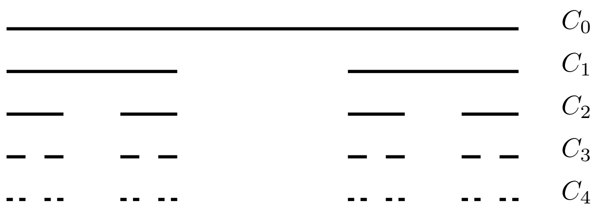
The outer Lebesgue measure in $\mathbb{R}^n$. For $\mathbb{R}^n$ instead of $\mathbb{R}$, intervals in the definition of λ^* are replaced with open axis-parallel boxes (in other words, Cartesian products of open intervals). Then it can be shown, with some more effort, that appropriate analogs of the statements above hold, i.e., boxes have the right outer measure and λ^* is countably subadditive.

The Banach–Tarski paradox. In dimension 3 and higher, the lack of finite additivity of the outer measure λ^* manifests itself in a particularly bizzare way. Namely, it can be shown, with fairly nontrivial proof, that a 3-dimensional ball B^3 can be decomposed into finitely many disjoint subsets $A_1, \ldots, A_k$ so that, after applying a suitable rotation and translation to each A_i, the resulting sets are again disjoint but they fill *two* disjoint copies of B^3. One says that B^3 and the two copies of B^3 are *equidecomposable*. In this particular case, it suffices to take $k = 5$.

Interestingly, a 2-dimensional analog of this paradox is impossible: two equidecomposable planar geometric figures must have the

same area. A standard treatment of the Banach–Tarski paradox and related problems is [**Wag93**], but many new results have been obtained since its publication, so readers interested in the current status of the various questions in [**Wag93**] should definitely consult the Internet.

Exercise 1.7. Let C be the Cantor "middle-third" set, which can be compactly defined as $\bigcap_{i=1}^{\infty} C_i$, where $C_0 = [0,1]$ and $C_i = \frac{1}{3}C_{i-1} \cup (\frac{1}{3}C_{i-1} + \frac{2}{3})$:



(a) Show that $\lambda^*(C) = 0$.

(b) Show that C is closed and uncountable. ✠

1.2. The Lebesgue Measure. The second step in the construction of the Lebesgue measure is defining a suitable system $\mathcal{E}$ of subsets of $\mathbb{R}$ (or $\mathbb{R}^n$, the definition is formally the same) such that the outer measure λ^* restricted to $\mathcal{E}$ becomes countably additive, and at the same time, $\mathcal{E}$ is as rich as possible. Then the Lebesgue measure λ we are after is the restriction of λ^* to this $\mathcal{E}$.

The complement of a set E is the set $\mathbb{R} \setminus E$. It will be denoted by E^c.

Definition 1.8. *We say that a set E is* **measurable** *if for every set $A \subseteq \mathbb{R}$ we have*

$$\lambda^*(A) = \lambda^*(A \cap E) + \lambda^*(A \cap E^c). \tag{1}$$

Let $\mathcal{E}$ be the system of all measurable subsets of $\mathbb{R}$. The **Lebesgue measure** *λ on $\mathbb{R}$ is the restriction of λ^* to $\mathcal{E}$.*

To check measurability of a set E, it suffices to verify the inequality $\geq$ in (1), since the opposite inequality is implied by the subadditivity of λ^*.

Exercise 1.9. Check that measurable sets can be approximated from the inside by compact sets. That is, for every $\varepsilon > 0$ and every measurable set E of finite measure there exists a compact set K such that $K \subseteq E$ and $\lambda^*(E \setminus K) < \varepsilon$. ✠

Likewise, measurable sets can be approximated from the outside by open sets. Using this, one can prove that a set E of finite outer measure is measurable if and only if for every $\varepsilon > 0$ there is a finite union F of open intervals such that $\lambda^*(F \Delta E) < \varepsilon$ (here $F \Delta E$ is the *symmetric difference* of the two sets, consisting of all points that belong to one of them but not the other one). That is, *every measurable set is nearly a finite union of intervals*, which is one of three informal observations known as *Littlewood's principles*.

We will now develop some properties of the system $\mathcal{E}$.

Observation 1.10. *If $\lambda^*(E) = 0$, then E is measurable.*

Proof. If A is an arbitrary set, then $A \cap E \subseteq E$ and $\lambda^*(A \cap E) \leq \lambda^*(E) = 0$. Similarly, $\lambda^*(A) \geq \lambda^*(A \cap E^c)$. Combining these two observations, we have

$$\lambda^*(A) \geq \lambda^*(A \cap E^c) = \lambda^*(A \cap E^c) + \lambda^*(A \cap E). \qquad \square$$

In the following exercise, diligent readers are asked to derive some basic properties of measurable sets. We recommend trying at least one of the parts, to gain some feeling for the definition of measurability.

Exercise 1.11. (a) Show that the interval $(0, \infty)$ is measurable.

(b) Prove that the complement of a measurable set is measurable.

(c) Verify that the union of two measurable sets is measurable, or more generally, that the union of countably many measurable sets is measurable. ✠

Set systems having the last two properties in the previous exercise are of fundamental importance in measure theory (and in probability theory), and they have a name:

Definition 1.12. *A* **$\boldsymbol{\sigma}$-algebra** *(some authors also use the term* σ-field*) is a nonempty set system closed under complementation and countable unions.*

Combining the two properties in the definition, it is easy to see that σ-algebras are also closed under countable intersections.

Exercise 1.13. Let X be a finite set. Describe all σ-algebras on X. ✠

We have already seen one example of a σ-algebra—the measurable sets on $\mathbb{R}$ or $\mathbb{R}^n$. If $\mathcal{A}$ is a system of subsets of a set X, the intersection $\mathcal{F}$ of all σ-algebras containing $\mathcal{A}$ is also a σ-algebra. We call it *the smallest σ-algebra containing* $\mathcal{A}$. We also say that $\mathcal{F}$ is the σ-algebra *generated by* $\mathcal{A}$.

Borel sets. If $\mathcal{A}$ is the family of all open intervals on $\mathbb{R}$, then the elements of the smallest σ-algebra containing $\mathcal{A}$ are called *Borel sets*.

To define Borel sets, we only needed basic set operations (to define σ-algebra) and the notion of open sets. These tools are available in any metric or topological space, and Borel sets can be defined there as well, in exactly the same way.

The notion of Borel sets appears quite often. Basically, when one considers a space with measure *and* metric (or topology), one almost always assumes that all Borel sets are measurable.

The family of Borel sets in $\mathbb{R}$ is a σ-algebra, and so, together with open intervals, it contains all of their countable unions. Likewise, it contains all countable intersections of these countable unions. But even if we iterate countable intersections and countable unions a thousand times, there will still be some Borel sets that we have not created. Borel sets can be defined inductively in a manner similar to what we have just described, but in this case the induction needs to be *transfinite*—we need ω_1 steps, where ω_1 is the first uncountable ordinal number.

Exercise 1.14. Prove that every Borel set in $\mathbb{R}$ is Lebesgue measurable. ✠

The converse is far from true; see Exercise 1.18.

It remains to prove that the Lebesgue measure is countably additive.

Theorem 1.15. *If $E_1, E_2, \ldots$ is a sequence of pairwise disjoint measurable sets, then*

$$\lambda\Big(\bigcup_{i=1}^{\infty} E_i\Big) = \sum_{i=1}^{\infty} \lambda(E_i).$$

Proof. Measurable sets form a σ-algebra, and so $\bigcup E_i$ is a measurable set.

We have already proved subadditivity, even for λ^*.

Let $E_1, E_2, \ldots$ be disjoint measurable sets. We take $A = E_1 \cup E_2$ as the "testing" set in the definition of a measurable set. Since E_1 is measurable, this definition tells us that

$$\lambda(E_1 \cup E_2) = \lambda(E_1) + \lambda(E_2).$$

This, of course, can be extended to any finite family of disjoint measurable sets.

Let us now suppose that the sequence $E_1, E_2, \ldots$ is countably infinite. Then

$$\lambda\Big(\bigcup_{i=1}^{\infty} E_i\Big) \geq \lambda\Big(\bigcup_{i=1}^{n} E_i\Big) = \sum_{i=1}^{n} \lambda(E_i)$$

for every natural number n, yielding $\lambda(\bigcup_{i=1}^{\infty} E_i) \geq \sum_{i=1}^{\infty} \lambda(E_i)$. □

On non-measurable sets. Probably the simplest known non-measurable set is the set V in Example 1.5. In order to define it, we needed to select one point from each of uncountably many equivalence classes, and this relies on the *axiom of choice* in set theory.

It is known that one cannot prove the existence of a non-measurable set without the axiom of choice. More precisely, the system of axioms of the usual (Zermelo–Fraenkel) set theory without the axiom of choice does not contradict the assumption "all subsets of $\mathbb{R}$ are Lebesgue measurable."

Since the axiom of choice is inherently nonconstructive, there is no way of fully describing a specific non-measurable set. You will never

see non-measurable sets in the daylight, but they may be lurking in the darkness—and you must protect your theorems against them.

Null sets and "almost everywhere." A set of measure zero is also often called a **null set**. One of the most often used properties of measure is that the union of finitely or countably many null sets is again a null set.

Here is a simple but typical way of using this. It is easy to check that a line in $\mathbb{R}^2$ has the (2-dimensional) Lebesgue measure 0. Hence the plane cannot be covered by countably many lines.

This gives us a very simple proof of existence of point sets in general position (no three collinear): having already constructed an n-point set P in general position, the points of P span finitely many lines, these do not cover $\mathbb{R}^2$, and so we can always add a new point to P while keeping general position. (Using transfinite induction, we can even obtain an uncountable set in general position in this way.) Similarly we can require no four of our points co-circular (since a circle has measure 0), and so on.

Here is another very common and useful piece of terminology: if Π is some property of a point in $\mathbb{R}^n$, and the set of points where Π does not hold is a null set, we say that Π holds **almost everywhere** (or for *almost all points*, etc.). Similarly, if $E \subset \mathbb{R}^n$ is a measurable set, we may say that Π holds almost everywhere on E. We will practice this terminology in the next remark.

A measurable set cannot be completely gray. Can we imagine what a measurable set looks like? One remarkable feature is that, on a sufficiently fine scale, it must be "grained"—it cannot be "gray" everywhere, i.e., to occupy half of each interval, say.

Let $E \subset \mathbb{R}^n$ be a measurable set. For a point $x \in \mathbb{R}^n$ we consider the limit

$$d_E(x) := \lim_{\delta \to 0} \frac{\lambda(B(x,\delta) \cap E)}{\lambda(B(x,\delta))},$$

where $B(x,r)$ denotes the (Euclidean) ball of radius r centered at x. The limit may not exist in general, but if it does, it is called the **density** of E at x.

The non-existence of a "completely gray" measurable set is expressed in the *Lebesgue density theorem*, asserting that every measurable $E \subseteq \mathbb{R}^n$ has density 1 at almost all of its points. That is, the set of those $x \in E$ for which $d_E(x)$ is undefined or smaller than 1 has Lebesgue measure 0.

Exercise 1.16. Prove that if $E \subseteq [0, 1]$ is measurable with $\lambda(E) > 0$, then $E \cap V$ is not measurable, where V is as in Example 1.5. ✠

Exercise 1.17. (a) Consider a modified Cantor set (sometimes called the Smith–Volterra–Cantor set) $\tilde{C} = \bigcap_{i=1}^{\infty} \tilde{C}_i$, (cf. Exercise 1.7), where the deleted intervals gradually occupy a smaller and smaller proportion of the remaining part: We set $\tilde{C}_1 = [0, 1]$ as before, and in the i-th step, $i = 1, 2, \ldots$, we have $\tilde{C}_i$ consisting of 2^{i-1} intervals of equal length. To obtain $\tilde{C}_{i+1}$, we remove the middle part of length 2^{-2i} from each interval in $\tilde{C}_i$. Show that $\tilde{C}$ is measurable and $\lambda(\tilde{C}) > 0$.

(b) Check that the characteristic function $\chi_{\tilde{C}}$ of $\tilde{C}$ is not Riemann integrable.

(c) Construct a sequence of continuous, and hence Riemann integrable, functions that converge to $\chi_{\tilde{C}}$. ✠

Exercise 1.18. (a) Construct a homeomorphism (a continuous map with continuous inverse) $\varphi\colon [0, 1] \to [0, 1]$ that maps the modified Cantor set $\tilde{C}$ from Exercise 1.17 to the "usual" Cantor set C (Exercise 1.7). Thus, homeomorphisms need not preserve null sets in general.

(b) Use (a) and Exercise 1.16 to exhibit a non-Borel measurable set. Hint: all subsets of the Cantor set are measurable. ✠

1.3. Measure Spaces and Important Examples. We started with $\mathbb{R}$, chose a particular family of subsets (which we called *measurable sets*), and defined a mapping from this family to extended real numbers (the Lebesgue measure). Albeit very important, this is just an instance of a more general concept.

Definition 1.19. *Let $\mathcal{F}$ be a σ-algebra. A* **measure** *on $\mathcal{F}$ is a function $\mu : \mathcal{F} \to \overline{\mathbb{R}}$ that is*

- *nonnegative,*
- *countably additive (if $E_1, E_2, \ldots$ are pairwise disjoint, then $\mu\left(\bigcup_j E_j\right) = \sum_j \mu(E_j)$), and*
- *assigns the value 0 to $\emptyset$.*

A **measure space** *is a triple $(X, \mathcal{F}, \mu)$, where X is a set, $\mathcal{F}$ is a σ-algebra on X, and μ is a measure on $\mathcal{F}$.*

As was the case with Lebesgue measure, $+\infty$ is an admissible value of μ. The elements of the σ-algebra $\mathcal{F}$ are called *measurable sets*. Usually the σ-algebra that we are referring to is clear from the context.

If we want to stress that we are talking about sets of real numbers satisfying Definition 1.8 (for which we have reserved the name measurable until now), we call them *Lebesgue measurable*.

If a property holds everywhere except for a set A with $\mu(A) = 0$, we say that it holds *μ-almost everywhere*.

Exercise 1.20. Let $A_1 \supseteq A_2 \supseteq \cdots$ be a nonincreasing sequence of μ-measurable sets with $\bigcap_{n=1}^{\infty} A_n = \emptyset$.

(a) Find an example, in which μ is the Lebesgue measure on $\mathbb{R}$, such that $\lim_{n\to\infty} \mu(A_n) \neq 0$.

(b) Now suppose that $\mu(A_1) < \infty$, and prove $\lim_{n\to\infty} \mu(A_n) = 0$. ✠

We proceed to important examples.

Counting and Dirac measures. If X is a set and $\mathcal{F}$ is the family of all subsets of X, the *counting measure* assigns to each set its number of elements. This measure does not distinguish infinite cardinalities, infinite sets are simply assigned $+\infty$.

Now let us fix a point $x_0 \in X$, and let $\mathcal{F}$ consist of all subsets of X again. The *Dirac measure* δ_{x_0} assigns 0 or 1 to each $A \in \mathcal{F}$, depending on whether $x_0 \in A$.

Hausdorff measure and Hausdorff dimension. The Lebesgue measure in $\mathbb{R}^n$ measures n-dimensional volume, but what if we want to measure the area of a surface in $\mathbb{R}^3$, for example?

A very general tool for measuring d-dimensional volumes in $\mathbb{R}^n$ is the d-dimensional **Hausdorff measure** H^d. Similar to the Lebesgue measure, the definition of H^d has two steps: first we define an outer measure H^{d*} on $\mathbb{R}^n$, and then we restrict it to an appropriate σ-algebra of measurable sets, where measurability is defined as in Definition 1.8, but with H^{d*} instead of λ^*.

The interesting part is the definition of H^{d*}, which can actually be used not only for subsets of $\mathbb{R}^n$, but for subsets of an arbitrary metric space.

First we define the **diameter** $\operatorname{diam} U$ of a set U as the supremum of distances of points $x, y \in U$. Next, for a set A and a real number $\delta > 0$, we set

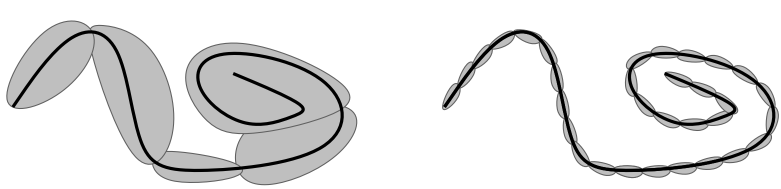
$$H^d_\delta(A) := \inf\left\{ \sum_{i=1}^{\infty} (\operatorname{diam} U_i)^d : \bigcup_{i=1}^{\infty} U_i \supseteq A, \operatorname{diam} U_i < \delta \text{ for all } i \right\}.$$

Finally,

$$H^{d*}(A) := \lim_{\delta \to 0} H^d_\delta(A)$$

(the limit, possibly infinite, exists since $H^d_\delta(A)$ is obviously nondecreasing as δ decreases).

To see how this works, let us first see how H^1 measures the length of a planar curve. For a given $\delta > 0$, we cover the curve with at most countably many sets of diameter smaller than δ, and we add up their diameters.



For some δ we may get much less than the length of the curve, but as δ gets smaller, the covering sets are forced to trace the curve more

and more closely, and it is at least intuitively plausible that we get a good notion of length.

Next, let us look at H^2 in the plane; then we are summing *squared* diameters. Let us consider the unit square Q, for example. Given $\delta > 0$, we can cover Q with $O(\delta^{-2})$ disks of diameter δ, and so $H^2_\delta(Q) = O(1)$.

On the other hand, we do need at least const $\cdot\, \delta^{-2}$ of sets of diameter at most δ in any covering: Choose an $n \times n$ square grid in Q. Then every two points are at least $\frac{1}{n}$ apart, and so each set of diameter less than $\frac{1}{n}$ can cover at most one grid point.

Hence $H^2(Q)$ is some positive constant. Actually, it turns out that in $\mathbb{R}^n$, the n-dimensional Hausdorff measure H^n is a constant multiple of the Lebesgue measure,[2] where the constant factor is the volume of a ball of radius $\frac{1}{2}$.

It may seem that the "most economic" covering should always be with balls of diameter δ. However, this is not the case in general: it can be shown that if we restrict only to coverings with balls, the measure we get is different from H^d for some sets.

An interesting feature is that d, the dimension, in the definition of H^d need not be an integer—the definition makes sense for every real $d \geq 0$. It is easy to check that $H^d \geq H^{d'}$ for $d \leq d'$. It turns out that for every set A there is exactly one "right" value d_0 of d, in the sense that for $d < d_0$ we have $H^d(A) = \infty$ and for $d > d_0$ we have $H^d(A) = 0$. (For d_0 itself, $H^{d_0}(A)$ can be a positive number, but also 0 or ∞.) This d_0 is called the **Hausdorff dimension** of A.

For example, the Cantor set C in Exercise 1.7 has Hausdorff dimension $\log_3 2 \approx 0.631$. Indeed, for every k, we can cover C by 2^k intervals of length 3^{-k}, and we cannot do much better. Most likely the reader has heard about *fractals*; the Cantor set is a very simple example of this popular concept. The Hausdorff dimension is very useful for measuring fractals, but also various exceptional sets in results of mathematical analysis, for example. (There are many other concepts of dimension, e.g., several topological ones, but these are usually only integer-valued.)

[2]Some authors re-normalize the Hausdorff measure so that H^n in $\mathbb{R}^n$ is the same as the Lebesgue measure.

Measuring the unit sphere. Quite often one needs a rotation-invariant measure μ on the unit sphere S^n in $\mathbb{R}^{n+1}$. The n-dimensional Hausdorff measure will do, but a more convenient and elementary definition is this: for a set $A \subseteq S^n$, define $\tilde{A}$ as the union of all segments connecting points of A to the center of the sphere, and define $\mu(A) := \lambda(\tilde{A})/\lambda(B^{n+1})$, where $B^{n+1} = \widetilde{S^n}$ is the unit ball.

The Haar measure. There are many ways of choosing a measure on an arbitrary set X. But a very useful fact of mathematics is that if X is a *group*, then, under reasonable assumptions, there is an essentially unique, canonical way of choosing a measure on X that is compatible with the group structure in a suitable sense. This result is known as *Haar's theorem*, and the respective measure as the *Haar measure*.

One thing that we take for granted when dealing with areas in the plane or volumes in $\mathbb{R}^n$ is that the volume of a set does not change if we translate the set. The Lebesgue measure satisfies this requirement—if the measure of a set $S \subset \mathbb{R}^n$ is defined, it is the same as the measure of the set $c+S$, for any $c \in \mathbb{R}^n$. We say that Lebesgue measure is invariant under the action of the group $\mathbb{R}^n$, with addition as the group operation.

Of course, there is nothing special about the additive group of reals (or $\mathbb{R}^n$). We would like to extend this concept to other (suitable) groups. This is the idea behind Haar measure. The Lebesgue measure on $\mathbb{R}^n$ is actually the completion of the Haar measure on $\mathbb{R}^n$ with the group operation of vector addition.

As another tiny example, let us consider the multiplicative group of positive real numbers. Clearly, Lebesgue measure is not invariant with respect to this group. But we can define a measure on the set of positive real numbers that does have this property, as follows:

$$\mu(S) := \int_S \frac{1}{x}\,\mathrm{d}x.$$

It is easy to check that whenever $0 < a < b$ and c is a positive real number, we indeed have $\mu([a,b]) = \mu([ca, cb])$.

One of the most important types of groups in all mathematics (and physics) are *matrix groups*; these are subgroups of the group $\mathrm{GL}(n, \mathbb{K})$ of all $n \times n$ invertible matrices over a field $\mathbb{K}$.

Let us consider the case $\mathbb{K} = \mathbb{R}$. Then $\mathrm{GL}(n, \mathbb{R})$ can be regarded as a subset of $\mathbb{R}^{n^2}$, which defines a metric (and topology) on $\mathrm{GL}(n, \mathbb{R})$, and allows us to speak of Borel sets in $\mathrm{GL}(n, \mathbb{R})$.

In contrast to the simple cases of additive or multiplicative groups of reals, we now need to start distinguishing left- and right-invariance. Haar's theorem implies that there is a measure ν on $\mathrm{GL}(n, \mathbb{R})$ that is *left translation-invariant* under matrix multiplication; that is, for every Borel set $S \subseteq \mathrm{GL}(n, \mathbb{R})$ and every matrix $A \in \mathrm{GL}(n, \mathbb{R})$, we have

$$\nu(AS) = \nu(S), \text{ where } AS = \{AB : B \in S\}.$$

Moreover, if we require certain reasonable properties of ν (finiteness on compact sets and regularity; see below), then ν is a *unique* left translation-invariant Borel measure on $\mathrm{GL}(n, \mathbb{R})$—more precisely, unique up to a scalar multiple (since if ν works, then 10ν works, too ...).

This ν is also right translation-invariant, $\nu(SA) = \nu(S)$; this is a special property of $\mathrm{GL}(n, \mathbb{R})$, not shared by all groups covered by Haar's theorem.

Let us now consider the subgroup $\mathrm{SO}(n, \mathbb{R}) \subseteq \mathrm{GL}(n, \mathbb{R})$ of all rotations in $\mathbb{R}^n$ (orthogonal matrices with determinant 1, that is). Then, again, there is a unique, up to scalar multiple, reasonable left translation-invariant Borel measure ν' on $\mathrm{SO}(n, \mathbb{R})$. It again happens to be right translation-invariant as well. Let us also note that ν' is *not* the restriction of the ν from above (since it can be shown that $\nu(\mathrm{SO}(n, \mathbb{R})) = 0$). While $\nu(\mathrm{GL}(n, \mathbb{R})) = \infty$, we have $\nu'(\mathrm{SO}(n, \mathbb{R}))$ finite.

Let us mention in passing that this ν' can also be used to define a rotation-invariant measure on the sphere: for a Borel set $E \subseteq S^{n-1}$, we set $\mu(E) := \nu'\{A \in \mathrm{SO}(n, \mathbb{R}) : Ax_0 \in E\}$, where $x_0 \in S^{n-1}$ is a fixed point. It can be shown that, up to scalar multiple, this is the same measure on S^{n-1} as the one we have introduced earlier in a more pedestrian way.

Stating Haar's theorem in general requires introducing several notions. This part is more technical than the rest of the chapter and a less experienced reader may skip it. The main message is that nice enough groups possess essentially unique reasonable left translation-invariant measure.

Haar's theorem is concerned with the situation where the group G in question is a **topological group**. This means that G is also a topological space[3] such that the group operations, multiplication and inversion, are continuous. Moreover, we want G to be a *locally compact topological space*, which in the case of a topological group means that the unit element $e \in G$ has a compact neighborhood (in the metric case, a sufficiently small closed ball around e is compact). We also need that G be *Hausdorff*, which is a fairly standard condition which holds for all metric spaces, for example, and only very exotic topological spaces fail to satisfy it. All of these conditions are satisfied for the matrix groups mentioned above.

The definition of a **regular measure** has two parts. A measure μ is *outer-regular on Borel sets* if $\mu(E) = \inf\{\mu(U) : U \supseteq E, U \text{ open}\}$ for every Borel set E (i.e., Borel sets can be approximated from outside by open sets), while μ is *inner-regular on open sets* if $\mu(U) = \sup\{\mu(K) : K \subseteq U, K \text{ compact}\}$ for every open set U. A *regular* measure is one that is both outer-regular on Borel sets and inner-regular on open sets.

Now we can state Haar's theorem (without proof):

Theorem 1.21 (Haar's theorem)**.** *Let G be a Hausdorff and locally compact topological group. Then there is a measure μ on the σ-algebra of Borel sets in G, unique up to a multiplicative constant, with the following properties:*

- *μ is nontrivial (attains some values in $(0, \infty)$);*
- *μ is left translation-invariant;*
- *μ is finite on all compact subsets of G;*
- *μ is regular in the sense above.*

[3]Readers not familiar with topology can think of a metric space, or a subspace of some $\mathbb{R}^n$ or possibly look into Chapter 6.

A μ as in the theorem is called the **left Haar measure**. One can define right Haar measure analogously. For compact G or commutative G, and in many other cases of interest, the left and right Haar measures coincide. However, for example for the group of all *invertible affine transformations of* $\mathbb{R}$, of the form $x \mapsto ax + b$, $a \neq 0$, with the operation of composition, the left and right Haar measures are not the same.

For the usual matrix groups, explicit formulas for the Haar measure are known, but often one needs just the existence (and sometimes uniqueness).

In addition to the examples already given, we mention that the Haar measure on $\mathrm{GL}(n, \mathbb{R})$ is given by

$$\nu(E) = \int_E \frac{1}{|\det(X)|^n} \,\mathrm{d}X$$

(Lebesgue integral in $\mathbb{R}^{n^2}$; see below).

2. The Lebesgue Integral

2.1. Measurable Functions. We are getting close to defining the Lebesgue integral of a function f. In many contexts (probability theory etc.), it is useful to do it in a general measure space $(X, \mathcal{F}, \mu)$, not just in $\mathbb{R}$. Unfortunately, we cannot define the integral for *all* functions, but only for *measurable* ones. Measurable functions preserve the structure of measure spaces, much like continuous functions preserve the structure of topological spaces (the preimage of an open set is an open set).

Definition 2.1. *Let X and Y be sets equipped with σ-algebras $\mathcal{F}$ and $\mathcal{G}$, respectively. A function $f\colon X \to Y$ is* **measurable** *if $f^{-1}(A) \in \mathcal{F}$ for every $A \in \mathcal{G}$.*

The next exercise provides a useful criterion of measurability.

Exercise 2.2. If $\mathcal{G}$ in Definition 2.1 is the smallest σ-algebra containing a set system $\mathcal{H}$, then, to show that f is measurable, one only needs to verify that $f^{-1}(A) \in \mathcal{F}$ for every $A \in \mathcal{H}$. Prove that. ✠

When both $X = Y = \mathbb{R}$, the usual convention (somewhat asymmetric!) is that $\mathcal{F}$ is the σ-algebra of Lebesgue measurable sets as defined in Section 1.2 and $\mathcal{G}$ are the Borel sets. Using Exercise 2.2, we can rewrite the definition as follows.

Definition 2.3. *Let $D \in \mathcal{F}$. A function $f\colon D \to \overline{\mathbb{R}}$ is* **measurable** *if the set $\{x \in D : f(x) \in I\}$ is measurable for every interval $I \subseteq \overline{\mathbb{R}}$.*

In Section 1.2 we mentioned the first of three *Littlewood's principles*. The other two have to do with measurable functions. They say that every measurable function is nearly continuous, and that every convergent sequence of measurable functions is nearly uniformly convergent. We will not formalize these observations here, but even in this form they are useful to provide some intuition about measurable sets and functions.

For a real function f we define its **positive part** f^+ as $\max\{f, 0\}$ and its **negative part** f^- as $\max\{-f, 0\}$. Besides the usual conventions regarding adding infinity to a real number and multiplying infinity by a real number, we define

$$0 \cdot (\pm\infty) = 0.$$

The expressions $\infty - \infty$, $\pm\infty/0$, $a/0$ (for real a) and $\pm\infty/\pm\infty$ remain undefined.

The following properties of measurable functions are easy to verify. If $f, f_1, f_2, \dots$ are measurable extended-real valued functions on $D \in \mathcal{F}$, then $|f|$, f^+, f^-, $\sup f_j$, $\inf f_j$, $\limsup f_j$, $\liminf f_j$ are measurable on D, $1/f$ is measurable on $\{x \in D; f(x) \neq 0\}$, and $f_1 + f_2$, $f_1 - f_2$, $f_1 f_2$, f_1/f_2 are measurable wherever these expressions are defined. Moreover, the set D' of points where $\lim f_j$ exists is measurable, and $f = \lim f_j$ is measurable on D'.

In the definition of the Lebesgue integral, *simple functions* play a role similar to the role of divisions of an interval in case of the Riemann integral.

Definition 2.4. *Let X be a set, $\mathcal{F}$ a family of subsets of X, and $D \in \mathcal{F}$. A function $f\colon D \to \mathbb{R}$ is* **simple** *if it is a linear combination of characteristic functions of sets in $\mathcal{F}$.*

The term *linear combination* implies that the sum is finite and the coefficients are real (no infinities). The definition makes sense for any set system $\mathcal{F}$, but the most important instance is when $\mathcal{F}$ is a σ-algebra. In this case a function is simple if and only if it is measurable and attains a finite number of values, all of them finite.

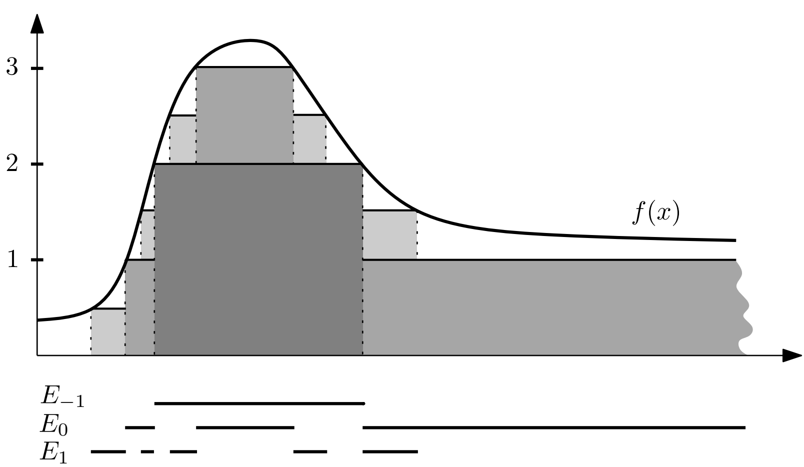
Proposition 2.5. *Let X be a set, $\mathcal{F}$ a σ-algebra on X, and f a nonnegative measurable function on $D \in \mathcal{F}$. Then one can find simple functions $f_1, f_2, \ldots$ that converge pointwise to f and such that $f_k \leq f_{k+1}$ for all k.*

Proof. We present a proof assuming that f attains only finite values. The extension so that values $+\infty$ are handled as well is not difficult and we leave it to the reader.

For $j \in \mathbb{Z}$ let us define $P_j = \bigcup_{i \text{ odd}} [i2^{-j}, (i+1)2^{-j})$ and $E_j = \{x \in D : f(x) \in P_j\}$. In other words, $x \in E_j$ if and only if 1 occupies the j-th position of the binary representation of $f(x)$. Since the P_j are Borel, the E_j are measurable. It is easy to see that

$$f = \sum_{j=-\infty}^{\infty} 2^{-j} \chi_{E_j}.$$

The functions $f_k = \sum_{j=-k}^{k} 2^{-j} \chi_{E_j}$ are simple and converge to f from below. Here is an illustration of this construction:



□

2.2. Lebesgue Integral, Convergence Theorems. Let $(X, \mathcal{F}, \mu)$ be a measure space. The Lebesgue integral of a function f on a set $D \in \mathcal{F}$ is usually denoted by $\int_D f \, \mathrm{d}\mu$. One can also encounter the notation $\int_D f$, $\int_D f(x) \, \mathrm{d}\mu(x)$, or even $\int_a^b f(x) \, \mathrm{d}x$ if D is the interval $[a, b]$.

In the last mentioned case, the notation is the same as the usual notation for the Riemann integral. However, it turns out that every Riemann integrable function on a closed interval is measurable, and that its Riemann and Lebesgue integrals coincide.

We will define the Lebesgue integral in three steps: for simple functions, then for nonnegative measurable functions, and finally for measurable functions with arbitrary signs.

Definition 2.6. *Let $(X, \mathcal{F}, \mu)$ be a measure space and $f = \sum_i a_i \chi_{A_i}$ a simple function on $D \in \mathcal{F}$. We define*

$$\int_D f d\mu := \sum_i a_i \mu(A_i).$$

There are many ways to write a given simple function. For instance, $\chi_{[0,2]} + \chi_{[1,3]}$ is the same function as $\chi_{[0,1]} + 2 \cdot \chi_{[1,2]} + \chi_{[2,3]}$. It turns out that the value of the integral does not depend on the way we write the function, but we skip the proof.

Definition 2.7. *Let $(X, \mathcal{F}, \mu)$ be a measure space and $D \in \mathcal{F}$.*

- *If f is a nonnegative measurable function on D, then*

$$\int_D f \, \mathrm{d}\mu = \sup \left\{ \int_D \psi \, \mathrm{d}\mu : \psi \textit{ is a simple function with } \psi \leq f \right\}.$$

- *If f is a measurable function on D, then $\int_D f = \int_D f^+ - \int_D f^-$, provided the difference makes sense (at least one of the integrals on the right side is finite). If the difference is $\infty - \infty$, the integral is not defined.*

Perhaps slightly confusingly, we call f *integrable* whenever the integral $\int_D f \, \mathrm{d}\mu$ not only exists, but, moreover, is finite.

Now the reason for considering only measurable functions becomes more clear. Recall that a function is Riemann integrable if the supremum of the vertical rectangles that fit under the curve is

equal to the infimum of the rectangles that approximate the curve from above. For the Lebesgue integral the situation is similar: for nonnegative functions f, we would like to see the equality

$$\inf_{f\le\psi}\int\psi=\sup_{\varphi\le f}\int\varphi$$

where ψ and φ are simple functions. If f is a bounded function on a set of finite measure, then the above equality is true if and only if there is a measurable function g that is equal to f almost everywhere.

Let us gather some basic properties of the Lebesgue integral. While the proofs are not difficult, proving at least some of them might help the reader to gain some insight and practice the definitions.

Proposition 2.8. *Let $D\in\mathcal{F}$ and let f,g be measurable functions on D.*

(i) *If f is nonnegative, $D_1,D_2\in\mathcal{F}$, and $D_1\subseteq D_2\subseteq D$, then $\int_{D_1}f\,\mathrm{d}\mu\le\int_{D_2}f\,\mathrm{d}\mu$.*
(ii) *If $D_1,D_2\in\mathcal{F}$, $D_1\cap D_2=\emptyset$, and $D_1\cup D_2=D$, then $\int_D f\,\mathrm{d}\mu=\int_{D_1}f\,\mathrm{d}\mu+\int_{D_2}f d\mu$.*
(iii) *If $\int_D|f|\,\mathrm{d}\mu<\infty$, then $|f|<\infty$ almost everywhere.*
(iv) *If $\int_D|f|\,\mathrm{d}\mu=0$, then $f=0$ almost everywhere.*
(v) *If f,g have integrals and $f\le g$ almost everywhere, then $\int_D f\,\mathrm{d}\mu\le\int_D g\,\mathrm{d}\mu$.*
(vi) *If $\int_D g\,\mathrm{d}\mu<\infty$ and $|f|\le g$ almost everywhere, then f is integrable.*

Exercise 2.9. Prove that if φ and ψ are nonnegative simple functions on $D\in\mathcal{F}$, then

$$\int_D\varphi+\psi=\int_D\varphi+\int_D\psi.$$ ✠

The Lebesgue integral is defined very generally and it is useful in a broad spectrum of situations. We do not impose many restrictions on the σ-algebra and the measure that we are using, but we pay for this by some loss of intuition. For instance, we saw in Exercise 2.9 that linearity of the integral is more or less obvious for simple functions. Linearity holds for general measurable functions as well,

but the proof is far from trivial. The easiest approach is probably through a monotone convergence theorem, as we will now show. We begin with another result, traditionally called a lemma.

Lemma 2.10 (Fatou's lemma). *Let $D \in \mathcal{F}$ and let $\{f_n\}_{n=1}^\infty$ be a sequence of nonnegative measurable functions defined on D, converging to a function f. Then*

$$\int_D f \leq \liminf_{n\to\infty} \int_D f_n.$$

Proof. It is enough to show that whenever φ is a nonnegative simple function satisfying $\varphi \leq f$, then $\int_D \varphi \,\mathrm{d}\mu \leq \liminf \int_D f_n \,\mathrm{d}\mu$.

If $\int_D \varphi \,\mathrm{d}\mu < \infty$, then the set $A = \{x \in D : \varphi(x) > 0\}$ is measurable and of finite measure. We choose an $\varepsilon > 0$ and let

$$A_n = \{x \in D : \text{ for all } k \geq n \text{ we have } f_k(x) \geq (1-\varepsilon)\varphi(x)\}.$$

The sequence $\{A_n\}$ is increasing, i.e., $A_n \subseteq A_{n+1}$ for all n, and $A \subseteq \bigcup A_n$. It follows that $\lim_{n\to\infty} \mu(A \setminus A_n) = 0$ (see Exercise 1.20), and there is an n such that $\mu(A \setminus A_k) < \varepsilon$ for all $k \geq n$. If M is the maximum of the function φ, then for $k \geq n$ we have

$$\begin{aligned}\int_D f_k &\geq \int_{A_k} f_k \geq (1-\varepsilon)\int_{A_k}\varphi = (1-\varepsilon)\int_D \varphi - (1-\varepsilon)\int_{A\setminus A_k}\varphi \\ &\geq (1-\varepsilon)\int_D \varphi - \int_{A\setminus A_k}\varphi \geq (1-\varepsilon)\int_D \varphi - \varepsilon M.\end{aligned}$$

But then $\liminf \int_D f_k \geq (1-\varepsilon)\int_D \varphi - \varepsilon M$. The choice of $\varepsilon > 0$ was arbitrary, and so $\liminf \int_D f_k \geq \int_D \varphi$.

The situation is similar when $\int_D \varphi = \infty$. In this case, we can find a number $a > 0$ such that the set $A = \{x \in D : \varphi(x) > a\}$ has an infinite measure. We set

$$A_n = \{x \in D : \text{ for all } k \geq n \text{ we have } f_k(x) \geq a\}.$$

The sequence A_n is again increasing, and its union covers A. A variation on the theme of Exercise 1.20, which we leave to the reader, shows that $\lim \mu(A_n) = \mu(A) = \infty$. Since $\int_D f_n \geq a\mu(A_n)$, we also have $\lim \int_D f_n = \infty = \int_D \varphi$. □

The conclusion of Fatou's lemma also holds with the weaker assumption that the f_n converge to f almost everywhere. The proof is almost identical—we just need to separate the set (of measure zero) on which the functions do not converge.

Here is one of the promised results stating that integration and limit can be interchanged under suitable conditions.

Theorem 2.11 (Monotone convergence theorem)**.** *Let $D \in \mathcal{F}$ and let $\{f_n\}$ be a sequence of nonnegative functions that converge to a function f almost everywhere on D. If, moreover, $f_n \leq f$ for all n, then*

$$\int_D f = \lim \int_D f_n.$$

Proof. Since $f_n \leq f$, we also have $\int_D f_n \leq \int_D f$. It follows that

$$\int_D f \leq \liminf \int_D f_n \leq \limsup \int_D f_n \leq \int_D f. \qquad \square$$

Now, finally, we can prove the linearity of the Lebesgue integral. Lemma 2.10 and Theorem 2.11 are important in their own right, and this is only one of their many applications.

Theorem 2.12. *If f and g are integrable measurable functions on $D \in \mathcal{F}$ and $c_1, c_2 \in \mathbb{R}$, then*

$$\int_D (c_1 f + c_2 g) = c_1 \int_D f + c_2 \int_D g.$$

Proof. It easily follows from the definition of integral that $\int_D cf = c \int_D f$.

Let f and g be nonnegative measurable functions. If $\{\varphi_n\}$ and $\{\psi_n\}$ are sequences of simple functions converging to f and g, respectively, then $\{\varphi_n + \psi_n\}$ is a sequence of simple functions converging to $f + g$. Assuming, as we may, $\varphi_n \leq f$ and $\psi_n \leq g$ for all n and using Theorem 2.11 and Exercise 2.9, we get

$$\int_D (f+g) = \int_D \lim(\varphi_n + \psi_n) = \lim \left(\int_D \varphi_n + \int_D \psi_n \right) = \int_D f + \int_D g.$$

If f and g are general measurable functions, we partition D into sets where both are nonnegative, both negative, etc., and we treat each of these sets separately. $\square$

The following theorem, due to Lebesgue, is another basic and very useful tool.

Theorem 2.13 (Dominated convergence theorem)**.** *Suppose that $D \in \mathcal{F}$, g is an integrable function on D, and $\{f_n\}$ is a sequence of measurable functions such that $|f_n| \leq g$ on D and the f_n converge to f almost everywhere on D. Then*

$$\int_D f = \lim \int_D f_n.$$

Proof. The functions $g - f_n$ are nonnegative, so we can apply Fatou's lemma and obtain

$$\int_D (g - f) \leq \liminf \int_D (g - f_n).$$

Since $|f| \leq g$, the function f is integrable and we have

$$\int_D g - \int_D f \leq \int_D g - \limsup \int_D f_n.$$

It follows that $\int_D f \geq \limsup \int_D f_n$. Starting with $g + f_n$ in place of $g - f_n$, we conclude that $\int_D f \leq \liminf \int_D f_n$. □

2.3. Product Measures, Fubini's Theorem. In combinatorics and counting, one of the most useful tricks is to interchange the order of summation, $\sum_i \sum_j = \sum_j \sum_i$. Fubini's theorem, treated in this section, provides an analogous tool for Lebesgue integration. For example, given a reasonable function $f\colon \mathbb{R}^2 \to [0, \infty)$, Fubini's theorem allows us to compute the volume between the graph of f and the xy plane by double integration: we can first integrate f with respect to the variable x, treating y as a constant, and then integrate the result with respect to y. Alternatively, we can first integrate with respect to y and then with respect to x.

As we will see, Fubini's theorem has several somewhat subtle assumptions. But as usual, it is much simpler and more general than an analogous result for the Riemann integral. We begin with some notions appearing in Fubini's theorem.

Definition 2.14. *Let $(X, \mathcal{F}, \mu)$ be a measure space. We say that μ is* **σ-finite** *if there is a sequence $X_1, X_2, \ldots$ of sets in $\mathcal{F}$ such that $\mu(X_j) < \infty$ for all j and $X = \bigcup_{j=1}^{\infty} X_j$.*

We say that μ is **complete** *if every subset of a set of measure zero is measurable.*

These two requirements ensure that our measure space is not overly exotic. For instance, $\mathbb{R}^n$ with the Lebesgue measure satisfies both of them.

Let $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{G}, \nu)$ be measure spaces such that the measures μ and ν are σ-finite. The elements of $\mathcal{F} \times \mathcal{G}$ are called *measurable rectangles.* For $A \in \mathcal{F}$ and $B \in \mathcal{G}$ we define

$$(\mu \times \nu)(A \times B) := \mu(A)\nu(B).$$

Then the **product measure** $\mu \otimes \nu$ is constructed in two steps: we first construct the corresponding outer measure $(\mu \times \nu)^*$, and then we define the σ-algebra $\overline{\mathcal{H}}$ of measurable sets with respect to $(\mu \times \nu)^*$, analogous to the way used for the Lebesgue measure (Definition 1.8). Then $\mu \otimes \nu$ is the restriction of $(\mu \times \nu)^*$ to $\overline{\mathcal{H}}$.

It can be shown that $\mu \otimes \nu$ is complete and coincides with $\mu \times \nu$ on all measurable rectangles, that $\overline{\mathcal{H}}$ is the smallest σ-algebra supporting a measure with these two properties, and that if both μ and ν are σ-finite, then $\mu \otimes \nu$ provides the only possible extension of $\mu \times \nu$ to the σ-algebra generated by the measurable rectangles.

We also remark that in many sources, the definition of the product measure is different: it lives on the σ-algebra generated by measurable rectangles, and it need not be complete. However, in this way, the product of the Lebesgue measure on $\mathbb{R}$ with itself is not the Lebesgue measure on $\mathbb{R}^2$, and so we prefer the definition above (which does guarantee that the $(n+m)$-dimensional Lebesgue measure is the product of the n- and m-dimensional Lebesgue measures).

Here is the main theorem of this section, which we state without proof.

Theorem 2.15 (Fubini's theorem)**.** *Suppose that* $(X, \mathcal{F}, \mu)$ *and* $(Y, \mathcal{G}, \nu)$ *are measure spaces with the measures* μ *and* ν *complete and* σ*-finite, and let* $(X \times Y, \mathcal{U}, \mu \otimes \nu)$ *be their (complete) product. Let* f *be a* $\mathcal{U}$*-measurable and integrable function on* $X \times Y$*. Then the following hold.*

(i) *For* μ*-almost all* x*, the function* f_x *defined by* $f_x(y) := f(x, y)$ *is an integrable function on* Y*. An analogous statement holds for* $f_y(x) := f(x, y)$*.*

(ii) $\int_Y f(x, y)\, \mathrm{d}\nu(y)$ *is an integrable function on* X*, and* $\int_X f(x, y)\, \mathrm{d}\mu(x)$ *is integrable on* Y*.*

(iii) $\int_X \left(\int_Y f\, \mathrm{d}\nu\right) \mathrm{d}\mu = \int_{X \times Y} f\, \mathrm{d}(\mu \otimes \nu) = \int_Y \left(\int_X f\, \mathrm{d}\mu\right) \mathrm{d}\nu.$

The same conclusion holds if we replace the requirement of integrability by the requirement that the function be nonnegative (this is known as *Tonelli's theorem*).

3. Foundations of Probability Theory

We assume that the reader has some background in probability; after all, it definitely belongs among the most important tools in other branches of mathematics and in science.

However, in order to introduce the basic notions of probability theory in a way that is both mathematically rigorous and sufficiently general, one needs the notion of measure—since probability *is* a kind of measure. Since measure is a relatively sophisticated notion, not so easily accessible to beginners, introductory courses on probability generally avoid laying proper foundations. Instead, they typically do finite probability spaces properly and geometric probability not at all, or in a semi-formal way.

Thus, having defined measure spaces, we want to point out their role in probability theory and state the usual axioms of probability.

3.1. Probability Spaces. What is a random point in the unit square? Everyone has some intuition about that; for example, imagine a square garden table, which is dry, but then it starts raining—the first drop hits the square at a random point.

In this case we are talking about the *uniform distribution* in the unit square, where no points are preferred to any others. It is tempting to say that every point should have the same probability, but this does not make much sense: there are infinitely many points in the square, and so the probability of hitting each particular one must be zero.

A reasonable thing to say is that for every geometric figure A in the square, the probability of hitting it should equal its area. So for geometric probability theory, we certainly need a definition of area. Moreover, if we want to conclude, for example, that hitting a point in the unit square with rational coordinates has zero probability, we are naturally led to countable additivity and thus to the notion of measure. However, the connection of contemporary probability theory to measure theory is much closer than just via geometric examples.

Axioms of probability. The current mathematical notion of probability took several hundred years to crystallize. It does not try to answer difficult philosophical questions like "What is randomness?", "Where does it come from?", "What is the meaning of probability in the real world?", etc., but it offers a mathematical model, which proves extremely successful in modeling real-world phenomena.

In most of the contemporary mathematical treatments of probability, the basic notion is a **probability space**, which is a mathematical model of some random process or experiment. As running examples, let us consider two simple experiments: three successive tosses of a fair coin (Example 3C), and picking a random point in the unit square (Example Sq).

A probability space is a triple $(\Omega, \mathcal{F}, \mathrm{P})$. The first component Ω is sometimes called the *sample space*, and it is a set consisting of all possible outcomes of the experiment. Each element $\omega \in \Omega$ is called an **elementary event**. For Example 3C, $\Omega_{3\mathrm{C}}$ consists of all possible three-letter sequences with letters H (heads) and T (tails): $\Omega = \{HHH, HHT, \ldots, TTT\}$. For Example Sq, $\Omega_{\mathrm{Sq}} = [0,1]^2$ consists of all points of the unit square.

The second component $\mathcal{F}$ of a probability space is a system of subsets of Ω. Each set $E \in \mathcal{F}$ is called an **event**. In Example 3C,

we admit *every* possible subset of Ω as an event, so $\mathcal{F}_{3C} = 2^{\Omega_{3C}}$. A concrete example of an event is "odd number of tails", which is the set $E = \{HHT, HTH, THH, TTT\}$. In Example Sq, the usual choice of $\mathcal{F}_{Sq}$ is the system of all Lebesgue measurable subsets of $[0,1]^2$: in this way, all reasonable geometric figures are events, and so we can talk about the probability of hitting the left half of the square, or hitting at most 0.1 from the center, etc., but "unreasonable" subsets of the square, i.e., sets that are not Lebesgue measurable, are not events.

The last component P of a probability space is a function that assigns a real number $\mathrm{P}(E)$, called the **probability of *E***,[4] to every event $E \in \mathcal{F}$. In Example 3C, we consider all of the elementary events equally likely, and the probability of an event E is defined as $\frac{|E|}{|\Omega_{3C}|}$. In Example Sq, we set $\mathrm{P}_{Sq}(E) = \lambda(E)$, the Lebesgue measure. (This works because the whole unit square has measure 1; if we considered a random point in some other geometric figure Ω in $\mathbb{R}^n$, we would need to take the ratio $\frac{\lambda(E)}{\lambda(\Omega)}$.)

The triple $(\Omega, \mathcal{F}, \mathrm{P})$ should satisfy the following axioms, which were first presented in this form by Kolmogorov in the 1930s:

Kolmogorov's axioms for probability space $(\Omega, \mathcal{F}, \mathrm{P})$

(0) The system $\mathcal{F}$ of events forms a σ-algebra. Explicitly, this means that Ω is an event, the complement of an event is an event, and so is a countable union of events.

(1) $\mathrm{P}(E) \geq 0$ for every $E \in \mathcal{F}$.

(2) $\mathrm{P}(\Omega) = 1$ (the experiment always has some outcome).

(3) P is countably additive: whenever $E_1, E_2, \ldots$ is a sequence of mutually disjoint events, then

$$\mathrm{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathrm{P}(E_i).$$

[4] In later chapters we will also be using $\mathrm{Prob}[E]$.

Axioms (1) and (2) are very intuitive, and so is finite additivity of probability. Countable additivity cannot really be substantiated by experience or intuition, but it becomes hard to avoid once we want to have reasonable notions of convergence of a sequence of random variables, which is a basic tool in probability theory and statistics.

A number of other natural properties of probability can be derived from the axioms, such as $\mathrm{P}(\emptyset) = 0$ or $\mathrm{P}(E) \leq \mathrm{P}(F)$ for $E \subseteq F$.

Comparing these axioms with the definition of a measure space, we find only one difference: the probability P is a measure with the additional condition $\mathrm{P}(\Omega) = 1$. Such a measure is called a **probability measure**.

A probability measure on a set Ω is also often referred to as a *probability distribution* on Ω.

Exercise 3.1. What is a probability space suitable for modeling the experiment "choosing three points a, b, c in the unit square, each of them uniformly distributed, the choices being mutually independent"? Hint: the triple needs to be represented by a single point in an appropriate higher-dimensional cube. ✠

Random real number and Benford's law. One mathematically simple but perhaps somewhat important remark is that there is no probability distribution on $\mathbb{R}$ in which "all real numbers have the same probability." If we wanted to make this vague phrase precise, we should require that the probability of hitting any interval I be proportional to its length. But any measure on $\mathbb{R}$ with this property must either be zero everywhere or assign measure ∞ to the whole $\mathbb{R}$, and so it cannot be a probability measure.

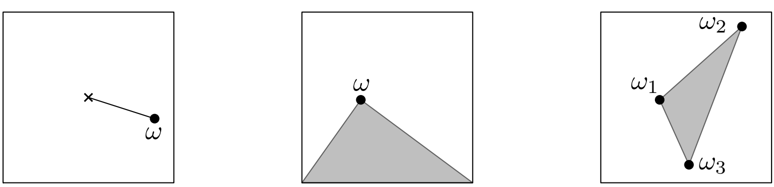
Yet in everyday life, in science, and at many other occasions we are confronted with various numbers that look quite random—your electricity bill, the area of the largest lake in your country, the number of followers of some Facebook group, etc.

We cannot resist mentioning a curious observation concerning such numbers. Namely, if we look at the leading digit of such "random" numbers, we encounter smaller digits considerably more often than larger ones: for example, 1 appears in about 30% cases, while 9 only about 5% of the time. It turns out that the probability of

digit i is proportional to $\log(\frac{i+1}{i})$. This is called *Benford's law*; apparently the phenomenon was first pointed out by Newcomb, who noticed that the first pages in tables of logarithms, those that contain numbers with leading digit 1, were more worn than the other pages. Benford's law is typically valid for distributions that span several orders of magnitude, and it has been used for detecting frauds in accounting and in science.

3.2. Random Variables and Expectation. The next key notion of probability theory, after a probability space, is a **random variable**. In the above setting, a (real) random variable X on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ is simply a P-measurable function $X : \Omega \to \mathbb{R}$ (random variables are typically denoted by capital letters). Explicitly, this means that $\{\omega \in \Omega : X(\omega) \le a\}$ is an event for every $a \in \mathbb{R}$ (or equivalently, $X^{-1}(B) \in \mathcal{F}$ for every Borel set $B \subseteq \mathbb{R}$).

Let us consider the probability space $(\Omega_{\mathrm{Sq}}, \mathcal{F}_{\mathrm{Sq}}, \mathrm{P}_{\mathrm{Sq}})$. Examples of random variables on it are the distance of a random point ω from the center of the square, or the area of the triangle spanned by ω and the two bottom corners. For the choice of three independent random points in the square as in Exercise 3.1, one may consider the area of the triangle spanned by the three points as a random variable.



The **expectation** $\mathbb{E}[X]$ of a random variable X is defined as the Lebesgue integral

$$\mathbb{E}[X] := \int_\Omega X(\omega)\, \mathrm{dP}(\omega).$$

Exercise 3.2. Compute the expectation of the first two random variables mentioned above. ✠

Alternative axiomatization: the algebra of random variables and free probability. Having talked about axioms of probability,

we should also mention a different approach, with very interesting generalizations, developed in the last few decades.

Its starting point is that for most problems in probability and statistics, there is a considerable freedom in choosing the underlying probability space, and the specific choice does not really matter—usually it suffices to know the expectations of various algebraic expressions in the random variables involved.

The alternative axiomatization of probability has random variables as elementary objects. They form an algebraic structure (called a *complex commutative* $*$*-algebra*), obeying suitable axioms, and there is a (linear) operator assigning to every random variable its expectation. When needed, an underlying probability space can then be constructed using suitable representation theorems from mathematical analysis.

This axiomatization has a generalization in which the random variables need not be commutative, in the sense that the expectation of XY may be different from the expectation of YX. In classical probability theory, such noncommutativity appears for matrix-valued random variables. This generalization leads to areas such as the theory of random matrices, quantum probability, or *free probability*, which is an active research field with connections to several other fields (including combinatorics), and which will almost surely be also useful for computer science. We refer, e.g., to [**Tao12**] for an introduction.

4. Literature

Measure and integration are basic and classical areas with plenty of textbooks. For example, Royden [**Roy88**] gives a friendly account, and Tao [**Tao11**] is a modern treatment with many interesting side views. Mattila [**Mat95**] nicely covers geometric aspects of measure theory, such as the construction and properties of the Hausdorff measure and its relatives. Even more textbooks exist for probability theory; here we mention Grimmett and Stirzaker [**GS01**].

Acknowledgment. We would like to thank Vojtěch Kaluža, Aleš Pultr, and Maria Saumell for reading and valuable comments.

Bibliography

[GS01] G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes*. Oxford University Press, New York, third edition, 2001.

[Mat95] P. Mattila. *Geometry of Sets and Measures in Euclidean Spaces*, volume 44 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995.

[Roy88] H. L. Royden. *Real Analysis*. Macmillan Publishing Company, New York, third edition, 1988.

[Tao11] T. Tao. *An Introduction to Measure Theory*, volume 126 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2011.

[Tao12] T. Tao. *Topics in Random Matrix Theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

[Wag93] S. Wagon. *The Banach–Tarski Paradox*. Cambridge University Press, Cambridge, 1993. With a foreword by Jan Mycielski, Corrected reprint of the 1985 original.

Chapter 2

High-Dimensional Geometry and Measure Concentration

In this chapter, we will discuss ideas and tools for dealing with geometric questions in high-dimensional spaces. We will encounter phenomena that are quite specific for large dimensions and contradict geometric intuition gained from the plane and 3-dimensional space.

Although our world is 3-dimensional, or perhaps 11-dimensional according to some physical theories, there are very good reasons for studying spaces of dimensions in the hundreds, millions, or more. Here are several settings where high-dimensional geometry comes into play.

- Results about discrete objects, such as graphs or partially ordered sets, have been proved by associating each object with a convex polytope, such as the *matching polytope* or *stable set polytope* for a graph or the *order polytope* for a partially ordered set, and then reasoning geometrically about it. The dimension of the polytope is typically comparable with the number of points of the discrete object; for example, it may be the number of vertices of the considered graph.
- Closely related to this, many computational tasks dealing with discrete objects are solved by continuous methods on

the associated polytopes. In particular, the *ellipsoid method* remains one of the most powerful theoretical tools for obtaining polynomial-time algorithms. Polyhedral optimization is often a method of choice for difficult practical problems. For example, train schedules in Holland were optimized using such tools already in the 1990s (while many other schedules remain sadly unoptimized up until now).

- A particularly impressive use of high-dimensional convex geometry can be seen at work in randomized algorithms for approximating the volume of convex bodies. These algorithms provide the only known method, e.g., for efficiently estimating the number of perfect matchings in a given graph.

- In recent years, efficient methods for processing large quantities of data have become particularly significant. Data sets need to be clustered, classified, learned, searched etc., and these are highly geometric problems, typically in huge dimensions.

  For example, a successful way of representing text documents is the *vector model*, where each possible word in the considered language(s) corresponds to one coordinate, and a document is represented by a vector whose coordinates specify the number of occurrences of each word. This representation, obviously, misses lot of information in the document since it ignores the order of the words, yet it retains enough to help locating documents on similar topics, or detecting plagiarism.

Links to probability. Modern high-dimensional geometry is inseparable from probability theory. Probabilistic thinking and terminology helps ones understanding of geometric questions, and geometry is helpful for achieving probabilistic results. Measuring volumes in the n-dimensional cube $[0,1]^n$ and reasoning about n independent random variables uniformly distributed in $[0,1]$ are two complementary and useful views of the same thing.

Here we assume that the reader knows basic notions of probability theory, like random variable, independence, and expectation, but we

do not rely on any sophisticated theorems from probability—we prove or at least recall almost everything we need. We primarily emphasize the geometric view, but we also try to offer a probabilistic perspective and to emphasize the correspondence of geometric and probabilistic notions.

Preliminaries. Throughout this chapter, for a set $A \subseteq \mathbb{R}^n$, $\lambda(A)$ denotes the n-dimensional Lebesgue measure of A. For simple geometric figures, such as balls or cubes, this is just the usual n-dimensional volume.

For $x \in \mathbb{R}^n$, $\|x\| = (x_1^2 + \cdots + x_n^2)^{1/2}$ denotes the Euclidean norm.

We assume that the reader is familiar with convex sets and functions, but, to be on the safe side, we recall the definitions here: A set $C \subseteq \mathbb{R}^n$ is *convex* if, for every $x, y \in C$, the segment $xy = \{(1-t)x + ty : t \in [0,1]\}$ is contained in C. A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is *convex* if $f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$ for all $x, y \in \mathbb{R}^n$ (more generally, a convex function may be defined only on a convex subset $C \subseteq \mathbb{R}^n$). A *concave* function satisfies the opposite inequality.

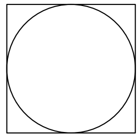
The *convex hull* $\operatorname{conv} X$ of a set $X \subseteq \mathbb{R}^n$ is the intersection of all convex sets in $\mathbb{R}^n$ containing X. A *convex polyhedron* is the intersection of finitely many halfspaces in $\mathbb{R}^n$. A *convex polytope* is a bounded convex polyhedron, or equivalently, the convex hull of a finite set of points. The term *convex body* is usually used for a convex set in $\mathbb{R}^n$ that is closed, bounded, and has nonempty interior (i.e., contains an n-dimensional ball).

1. Peculiarities of Large Dimensions

The ball in the cube. How can we generate a random direction in $\mathbb{R}^n$, i.e., a random point on the $(n-1)$-dimensional unit sphere S^{n-1}?

One way which comes to mind is to generate a random point z in the unit ball B^n bounded by S^{n-1}, and then project it to the sphere; in other words, we return $z/\|z\|$.

So how can we obtain a random point in B^n? Here is an idea. Consider the cube $[-1,1]^n$ circumscribed to B^n:



In the cube it is easy to generate a random point: we generate the n coordinates independently, each uniformly distributed in $[-1, 1]$. (Assuming that we have a generator of independent uniformly distributed random numbers, which is a highly nontrivial matter but here we leave it aside.)

To get a random point in B^n, we can use *rejection sampling*: Produce a random point in the cube and test if it lies in the ball. If yes, it is the desired point. If not, discard it and try again, until the ball is hit.

This looks like a reasonable method, and indeed it is if the dimension n is small. For example, for $n = 3$, the balls fills about 51% of the cube, and so it will be hit about every second time. But what about higher dimensions, say 20 or 100?

A formula for the volume of a ball can be found in many places, and there are several ways of calculating it. A particularly neat one will be indicated in Exercise 3.4 below. The formula reads

$$\lambda(B^n) = \frac{\pi^{\lfloor n/2 \rfloor} 2^{\lceil n/2 \rceil}}{\prod_{i:0 \leq 2i < n}(n - 2i)}.$$

By Stirling's formula for the factorial, one can show that $\lambda(B^n)$ behaves roughly like $(\frac{\pi e}{2n})^{n/2}$. This decreases exponentially as n grows, and for large n, the inscribed ball is like a negligible dust particle in the cube.

For $n = 11$, we expect to generate more than 1000 random points in the cube before hitting the ball, and for $n = 20$ about 40 million—so the rejection sampling method is rather useless in the latter case. We will see a better method for generating a random direction in Section 3.

Instead of the cube, one might try to approximate the ball by some other convex polytope. However, if we want to enclose B^n in a convex polytope P so that the ball fills, say, at least 1% of P, then the number of facets of P must be at least exponential in n (and hence

P cannot be defined as the intersection of less than exponentially many halfspaces); see Exercise 4.6.

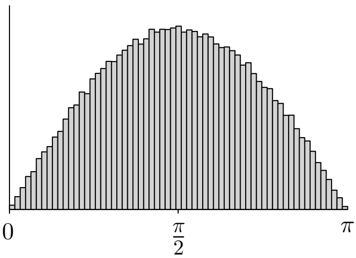
A similar but "dual" phenomenon is that, for large n, the convex hull of not too many points in B^n can fill only a negligible fraction of the ball, if the dimension is large. The next exercise provides a relatively simple proof, due to Elekes, of this fact, although quantitatively the bound can still be strengthened.

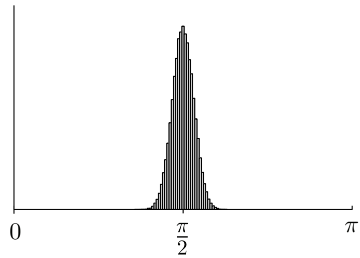
Exercise 1.1. Let $V \subset \mathbb{R}^n$ be a finite set. Prove that

$$\operatorname{conv} V \subseteq \bigcup_{v \in V} B(\tfrac{1}{2}v, \tfrac{1}{2}\|v\|),$$

where $B(x, r)$ is the ball of radius r centered at x. Deduce that the convex hull of N points contained in B^n has volume at most $\frac{N}{2^n}\lambda(B^n)$. ✠

Random unit vectors are almost perpendicular. Speaking about random directions in $\mathbb{R}^n$, here are distributions of the angle of 10^5 pairs of random unit vectors, in dimensions $n = 3$ and $n = 100$:





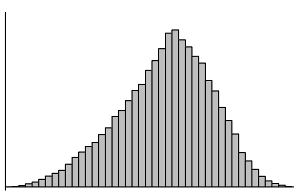
We can see that while all kinds of angles occur for $n = 3$, in dimension 100 the angles concentrate quite sharply around $\frac{\pi}{2}$.

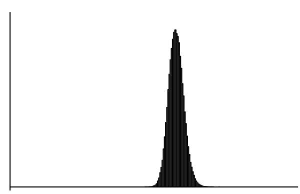
Here is a quick explanation of this phenomenon. Let us consider two random unit vectors $u, v \in S^{n-1}$. What can we say about the distribution of the inner product $\langle u, v\rangle$? By symmetry, we can assume that u is fixed, say to $e_1 = (1, 0, \ldots, 0)$. Then $\langle e_1, v\rangle = v_1$, the first coordinate. By symmetry again, the first coordinate of a random $v \in S^{n-1}$ has expectation 0. What is the variance $\operatorname{Var} v_1 = \mathbb{E}[v_1^2]$? Since $\sum_{i=1}^n v_i^2 = 1$ for unit v, we have $\sum_{i=1}^n \mathbb{E}[v_i^2] = 1$, and by symmetry for the third time, $\mathbb{E}[v_1^2] = \frac{1}{n}$.

Chebyshev's inequality tells us that, for all $t > 0$, a random variable X deviates from its expectation by $t\sqrt{\operatorname{Var} X}$ with probability at most $1/t^2$. So in our case, the absolute value of the inner product $\langle u, v\rangle$ of two random unit vectors exceeds $\frac{2}{\sqrt{n}}$ with probability at most $\frac{1}{4}$, for example.

We will see that the concentration is much stronger than what follows from Chebyshev's inequality: the probability of $|\langle u, v\rangle| \geq \frac{t}{\sqrt{n}}$ decreases *exponentially* in t, rather than just as $\frac{1}{t^2}$; see Exercise 4.7. This is a part of a much more general phenomenon: every "reasonable" function of a random variable from a "high-dimensional" space is strongly concentrated around its expected value. Later on, we will give the words "reasonable" and "high-dimensional" a precise meaning at least in some specific cases.

As another illustration, here is a plot of the distribution of the Euclidean distance of 10^5 random points in the n-dimensional unit cube $[0,1]^n$ from the center of the cube, again for $n = 3$ and $n = 100$.





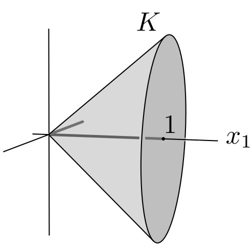
The high-dimensional picture looks similar to the one for angles (and indeed the limiting distributions for $n \to \infty$ turn out to be the same up to scaling).

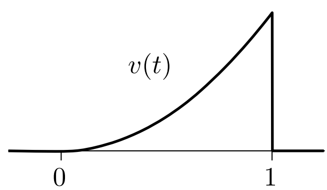
2. The Brunn–Minkowski Inequality and Euclidean Isoperimetry

Brunn's inequality for parallel slices. Let us slice a 3-dimensional convex loaf K of bread by three parallel planar cuts. Then the middle cut cannot have area smaller than both of the others: this is quite intuitive and it will follow from the Brunn–Minkowski inequality considered in this section.

If $v(t)$ denotes the area of the cut of K by the plane $x_1 = t$, then we claim that the function $v(t)$ is **unimodal**, meaning that for some t_0, it is nondecreasing on $(-\infty, t_0]$ and nonincreasing on $[t_0, \infty)$. A similar result holds for every convex body $K \subseteq \mathbb{R}^n$.

How can one prove unimodality? For a convex body K in the plane, the function $v(t)$ is easily checked to be concave on the interval $[t_{\min}, t_{\max}]$ of those t where the slices are nonempty, and every concave function is clearly unimodal. However, in dimension 3, concavity no longer holds: for example, if $K \subseteq \mathbb{R}^3$ is a rotational cone as in the picture (with opening angle 90 degrees),





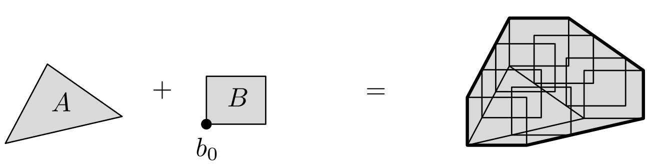
then $v(t) = \pi t^2$ for $t \in [0, 1]$, surely not concave.

The right function that is concave on $[t_{\min}, t_{\max}]$, for every n-dimensional convex body, is $v(t)^{1/(n-1)}$; this is a result of Brunn.

Minkowski sum and the Brunn–Minkowski inequality. Later Minkowski found a much more general result. To state it, we need the notion of a **Minkowski sum** of two sets $A, B \subseteq \mathbb{R}^n$, which is defined by

$$A + B := \{a + b : a \in A, b \in B\}.$$

One way of thinking of $A + B$ is as follows: fix a point $b_0 \in B$ and translate B so that b_0 sweeps all of A; then the union of these translates is $A + B$:



Exercise 2.1. True or false?

(a) The Minkowski sum of two Euclidean balls is a Euclidean ball.

(b) The Minkowski sum of two ellipses is an ellipse (here by an ellipse we mean the region, not only the curve). ✠

Exercise 2.2. Prove that if $A, B \subseteq \mathbb{R}^n$ are compact (i.e., closed and bounded), then so is $A + B$. ✠

Theorem 2.3 (Brunn–Minkowski inequality)**.** *Let A and B be nonempty compact sets in $\mathbb{R}^n$. Then*

$$\lambda(A+B)^{1/n} \geq \lambda(A)^{1/n} + \lambda(B)^{1/n}.$$

Assuming nonemptiness of A and B is important (why?). We also note that all compact sets are Lebesgue measurable, and the Minkowski sum of two compact sets is compact, and hence $\lambda(A)$, $\lambda(B)$, and $\lambda(A+B)$ are well defined.

We postpone the proof of this theorem to the next section.

Deriving Brunn's result. What does Theorem 2.3 have to do with Brunn's result for slice volumes?

Without loss of generality we may assume that the considered three parallel nonempty $(n-1)$-dimensional slices of the convex body $K \subset \mathbb{R}^n$ are at $x_1 = 0$, $x_1 = t$, and $x_1 = 1$, where $0 < t < 1$. If A is the left slice and B the right slice, then it is easy to see that, because of convexity of K, the middle slice contains the set $C := (1-t)A + tB$ (where $tB = \{tb : b \in B\}$ is B scaled down by t).

Now let $A' \subseteq \mathbb{R}^{n-1}$ be obtained from A by forgetting the x_1-coordinate of every point, and similarly for B' and C'. Then, by the Brunn–Minkowski inequality in $\mathbb{R}^{n-1}$, we have

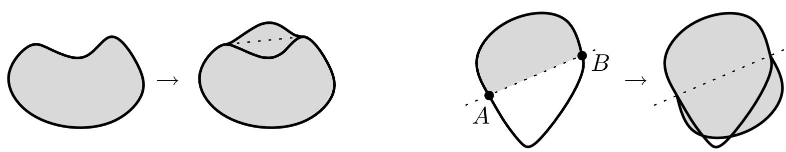
$$\begin{aligned} \lambda(C')^{1/(n-1)} &\geq \lambda((1-t)A')^{1/(n-1)} + \lambda(tB')^{1/(n-1)} \\ &= (1-t)\lambda(A')^{1/(n-1)} + t\lambda(B')^{1/(n-1)}. \end{aligned}$$

This gives the concavity of $v(t)^{1/(n-1)}$. □

2.1. Isoperimetry. The *isoperimetric problem* of ancient Greece asks for the largest area that can be fenced with a rope of given

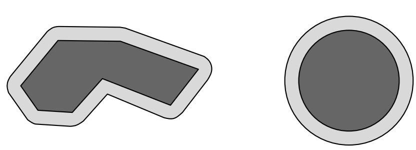
length. The Greeks knew that the answer should be a circular disk (and every physicist knows the answer to the analogous question in dimension 3, because a drop, whose surface is contracted by the surface tension, is spherical unless deformed by external force).

But the first rigorous proof was found only in the 19th century. Significant progress towards the solution was made by Steiner in 1838. Using the transformation on the left picture, he argued that a maximum-area geometric figure of a given perimeter must be convex,



and using the operation depicted on the right (where the points A and B divide the perimeter into two parts of equal length), he showed that such a figure must be a disk. However, the nontrivial step of showing that a maximum-area figure *exists* was still missing, and it was completed only later.

For a set $A \subseteq \mathbb{R}^n$ and a number $t > 0$, we define the ***t*-neighborhood** of A, denoted by A_t, as the set of all points with distance at most t to A. (Here we use the Euclidean distance, but the notion of t-neighborhood can be used in any metric space.) A modern formulation of the isoperimetric problem does not use the notion of perimeter (or surface area in higher dimensions); rather, it asks, for given $t > 0$, which set of a given volume has the smallest volume of the t-neighborhood.



In the picture, assuming that the dark areas are the same, the light gray area is the smallest for the disk.

The following result gives the expected answer, the ball, for Euclidean spaces, and it has a beautiful short proof using the Brunn–Minkowski inequality.

Theorem 2.4 (Euclidean isoperimetry). *For every measurable set* $A \subset \mathbb{R}^n$ *with* $\lambda(A) < \infty$ *and every* $t \geq 0$, *we have* $\lambda(A_t) \geq \lambda(B_t)$, *where* B *is a ball of volume* $\lambda(A)$.

Proof. For convenience we re-scale so that B is the ball of unit radius. Then $A_t = A + tB$, and so

$$\begin{aligned}\lambda(A_t) &= \lambda(A + tB) \geq \left(\lambda(A)^{1/n} + t\lambda(B)^{1/n}\right)^n \\ &= (1+t)^n \lambda(B) = \lambda(B_t).\end{aligned}$$

□

If we *define* the surface area of a set A as the limit

$$\lim_{t \to 0} \frac{1}{t}(\lambda(A_t) - \lambda(A))$$

(provided that the limit exists, that is), then the theorem just proved immediately yields that the ball has the smallest surface area among all sets of given measure for which this surface area is defined.[1]

The formulation of the isoperimetric problem using t-neighborhoods has the advantage that it uses only the notion of distance (metric) and volume (measure). Hence it makes sense in every metric measure space (i.e., a space with a metric and a measure, usually with the assumption that all balls are measurable). We will hear more about isoperimetric problems later in this chapter.

2.2. Proof of Brunn–Minkowski via Prékopa–Leindler. The Brunn–Minkowski inequality has several proofs. The one we will show is not the shortest or simplest known one, but its structure is in a sense typical for modern proofs of geometric inequalities.

First we are going to transform the Brunn–Minkowski inequality into a form in which the dimension n does not explicitly appear.

[1] This notion of surface area is called the *Minkowski content*. It is well defined for all convex bodies and many other sets, and for "reasonable" sets (polyhedra, sets with smooth boundary, etc.) it coincides with other definitions of surface area, such as the $(n-1)$-dimensional Hausdorff measure of the boundary. On the other hand, the Minkowski content is not a measure.

Then we regard it as an inequality about the characteristic functions of the sets involved, and we generalize this inequality to arbitrary nonnegative functions; this generalization is called the *Prékopa–Leindler inequality*. Most of the work in the proof of the Prékopa–Leindler inequality is in the 1-dimensional case, while the passage to higher dimensions is a straightforward induction.

As we go from the original inequality of Brunn for slice volumes to the Brunn–Minkowski inequality for arbitrary compact sets and further to the Prékopa–Leindler inequality for functions, geometric intuition gets lost and the formulas may start looking more and more arbitrary and daunting to a newcomer. But the advantage is increasing power and conceptually easier proofs.

Brunn–Minkowski on the line. The first step is establishing a one-dimensional Brunn–Minkowski inequality. For technical reasons we want it for arbitrary Lebesgue measurable sets, rather than just compact ones, but this makes almost no difference.

Lemma 2.5. *Let $A, B \subseteq \mathbb{R}$ be nonempty, measurable sets. Then $\lambda^*(A+B) \geq \lambda(A) + \lambda(B)$.*

Perhaps surprisingly, $A + B$ need not be measurable even if A and B are, so we use the outer measure λ^* in the formulation. Practically the inequality means that every measurable set containing $A+B$ has measure at least $\lambda(A) + \lambda(B)$.

Proof. First assume that A and B are compact, and translate them so that $\max A = 0 = \min B$. Then $A \cap B = \{0\}$, and $A + B \supseteq (A + \{0\}) \cup (\{0\} + B) = A \cup B$. Hence $\lambda(A + B) \geq \lambda(A \cup B) = \lambda(A)+\lambda(B)$. (We recall that $A+B$ is compact and hence measurable.)

For measurable A, B that are not necessarily compact, we use a fact from measure theory, namely, a Lebesgue measurable set A of measure $a < \infty$ contains, for every $\varepsilon > 0$, a compact subset A' of measure at least $a - \varepsilon$ (Exercise 1.9 in Chapter 1). By considering such compact subsets for both A and B, we get that $A + B$ contains a subset of measure at least $\lambda(A) + \lambda(B) - 2\varepsilon$ for every $\varepsilon > 0$, and the lemma follows. □

A dimension-free Brunn–Minkowski. Next, we reformulate the Brunn–Minkowski inequality as follows: for every compact $A, B \subset \mathbb{R}^n$ and every $t \in (0,1)$, we have

$$\lambda((1-t)A + tB) \geq \lambda(A)^{1-t}\lambda(B)^t. \tag{2}$$

One subtle advantage of this formulation is that here, unlike in Theorem 2.3, we no longer need to assume that A or B are nonempty.

Lemma 2.6. *The validity of* (2) *for all* A, B, t *implies the Brunn–Minkowski inequality in Theorem* 2.3*.*

Proof. If $\lambda(A) = 0$ or $\lambda(B) = 0$, then the desired inequality $\lambda(A + B)^{1/n} \geq \lambda(A)^{1/n} + \lambda(B)^{1/n}$ is obvious, so we assume $\lambda(A) > 0$ and $\lambda(B) > 0$ and we set

$$\alpha := \lambda(A)^{1/n}, \quad \beta := \lambda(B)^{1/n}, \quad t = \frac{\beta}{\alpha+\beta}, \quad A' := \frac{1}{\alpha}A, \quad B' := \frac{1}{\beta}B.$$

Apply (2) for A' and B': the right-hand side is 1, and the left-hand side equals $\lambda((1-t)A' + tB') = \lambda(\frac{\alpha}{\alpha+\beta}A' + \frac{\beta}{\alpha+\beta}B') = (\alpha+\beta)^{-n}\lambda(A+B)$. Thus $\lambda(A + B)^{1/n} \geq \alpha + \beta = \lambda(A)^{1/n} + \lambda(B)^{1/n}$. □

The Prékopa–Leindler inequality. If f, g, h are the characteristic functions of A, B, and $C := (1-t)A + tB$, respectively, we have $\lambda(A) = \int_{\mathbb{R}^n} f(x)\,\mathrm{d}x$, and similarly for B and C. Moreover, h is related to f and g as follows: $f(x) = 1$ and $g(y) = 1$ implies $h((1-t)x + ty) = 1$. Hence we also have $h((1-t)x + ty) \geq f(x)^{1-t}g(y)^t$ for all $x, y \in \mathbb{R}^n$, and this turns out to be the right assumption for the "functional form" of (2), where f, g, h are no longer characteristic functions of sets.

Theorem 2.7 (Prékopa–Leindler inequality)**.** *Let* $t \in (0,1)$ *and let* f, g, h *be functions* $\mathbb{R}^n \to \mathbb{R}$ *that are measurable, nonnegative, bounded, and have a finite integral over* $\mathbb{R}^n$*. Suppose that* $h\big((1-t)x + ty\big) \geq f(x)^{1-t}g(y)^t$ *for all* $x, y \in \mathbb{R}^n$*. Then*

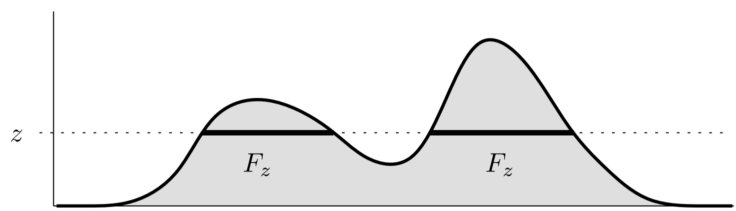
$$\int_{\mathbb{R}^n} h \geq \left(\int_{\mathbb{R}^n} f\right)^{1-t} \left(\int_{\mathbb{R}^n} g\right)^t.$$

By the preceding discussion, the Prékopa–Leindler inequality applied to the characteristic functions of the sets A, B, and $A+B$ implies the Brunn–Minkowski inequality.

Exercise 2.8. In order to familiarize yourself with the statement of the Prékopa–Leindler inequality, check that it suffices to prove it under the additional assumption that $\sup f = \sup g = 1$. ✠

Proof. As was announced above, the core of the proof is in dimension 1, and it relies on the 1-dimensional Brunn–Minkowski inequality (Lemma 2.5). We use the extra assumption $\sup f = \sup g = 1$ from the above exercise.

For $z \in [0,1]$, let $F_z := \{x \in \mathbb{R} : f(x) \geq z\}$ be the horizontal slice through the area below the graph of f at height z.



More or less from the definition of the Lebesgue integral, we have

$$\int_{\mathbb{R}} f(x)\,\mathrm{d}x = \int_0^1 \lambda(F_z)\,\mathrm{d}z.$$

With an analogous notation G_z and H_z, from the assumption $h((1-t)x+ty) \geq f(x)^{1-t} g(y)^t$ we obtain $H_z \supseteq (1-t)F_z + tG_z$ for every $z \in (0,1)$ (correct?). So the 1-dimensional Brunn–Minkowski inequality yields $\lambda(H_z) \geq (1-t)\lambda(F_z) + t\lambda(G_z)$. Note that we need F_z and G_z nonempty here, but this holds because of the assumption $\sup f = \sup g = 1$.

We will also need the weighted arithmetic-geometric mean inequality: for real numbers $a, b > 0$ and $t \in (0,1)$, we have $(1-t)a + tb \geq a^{1-t}b^t$.

Then we calculate

$$\begin{aligned}\int_{\mathbb{R}} h(x)\,\mathrm{d}x &= \int_0^1 \lambda(H_z)\,\mathrm{d}z\\ &\geq (1-t)\int_0^1 \lambda(F_z)\,\mathrm{d}z + t\int_0^1 \lambda(G_z)\,\mathrm{d}z\\ &\geq \left(\int_0^1 \lambda(F_z)\,\mathrm{d}z\right)^{1-t}\left(\int_0^1 \lambda(G_z)\,\mathrm{d}z\right)^t\\ &= \left(\int_{\mathbb{R}} f(x)\,\mathrm{d}x\right)^{1-t}\left(\int_{\mathbb{R}} g(x)\,\mathrm{d}x\right)^t.\end{aligned}$$

This finishes the 1-dimensional case, and we proceed with the induction on the dimension n.

So let $f, g, h\colon \mathbb{R}^n \to \mathbb{R}$ be as in Prékopa–Leindler. Not surprisingly, we want to apply the $(n-1)$-dimensional statement already proved. To get $(n-1)$-dimensional functions, we simply fix the last coordinate:

$$f_s(x_1,\dots,x_{n-1}) := f(x_1,\dots,x_{n-1},s),$$

and similarly for g_s and h_s. Now we claim that f_a, g_b, and h_c with $c := (1-t)a+tb$ satisfy the assumption of the $(n-1)$-dimensional Prékopa–Leindler. This is because $h_c((1-t)x+ty) = h((1-t)(x,a)+t(y,b)) \geq f(x,a)^{1-t}g(y,b)^t = f_a(x)^{1-t}g_b(y)^t$, where (x,a) is a shorthand for $(x_1,\dots,x_{n-1},a)$. So by the inductive hypothesis

$$\int_{\mathbb{R}^{n-1}} h_c \geq \left(\int_{\mathbb{R}^{n-1}} f_a\right)^{1-t}\left(\int_{\mathbb{R}^{n-1}} g_b\right)^t.$$

Now we let $F(a) := \int_{\mathbb{R}^{n-1}} f_a$ and similarly for $G(b)$ and $H(c)$. Then, magically, the last inequality is exactly the *assumption* of the 1-dimensional Prékopa–Leindler for F, G, H, and hence we obtain

$$\int_{\mathbb{R}} H \geq \left(\int_{\mathbb{R}} F\right)^{1-t}\left(\int_{\mathbb{R}} G\right)^t.$$

And, if we substitute the definitions of F, G, H and use Fubini's theorem, we obtain the conclusion of the n-dimensional Prékopa–Leindler inequality. □

Log-concavity. Speaking of the Prékopa–Leindler inequality, we cannot resist mentioning a related notion: a nonnegative function $f\colon \mathbb{R}^n \to \mathbb{R}$ is called **log-concave** if it satisfies

$$f((1-t)x+ty) \geq f(x)^{1-t} f(y)^t \text{ for all } x, y \in \mathbb{R}^n, t \in (0,1).$$

This is like the assumption of Prékopa–Leindler but with only one function.

If f is strictly positive, then log-concavity is the same as concavity of $\log f$, which explains the name.

The characteristic function of a convex set is log-concave, and so is the density of the Gaussian distribution discussed below. Many results and algorithms for convex sets generalize, sometimes with additional effort, to log-concave functions.

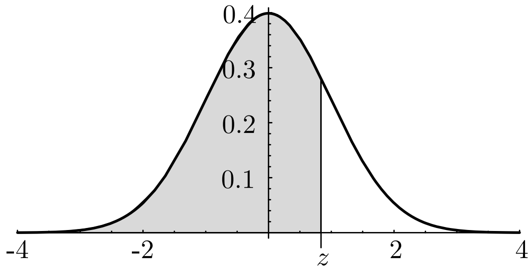
Exercise 2.9. Check that log-concave functions are closed under products and projections (or *marginals* in probability language); the latter means that if $f\colon \mathbb{R}^{m+n} \to \mathbb{R}$ is log-concave, then so is $g(x) = \int_{\mathbb{R}^n} f(x,y)\,\mathrm{d}y$. Hint: what was the main topic of this section? ✠

3. The Standard Normal Distribution and the Gaussian Measure

The **standard normal distribution** (or *Gaussian distribution*) on the real line has density function $\varphi(x) := ce^{-x^2/2}$, where $c > 0$ is a suitable constant—its value will be recalled soon. This means that for a random variable Z with the standard normal distribution, the probability of attaining a value not exceeding some z is given by the integral $\int_{-\infty}^{z} \varphi(x)\,\mathrm{d}x$; in symbols,

$$\mathrm{Prob}[Z \leq z] = \int_{-\infty}^{z} \varphi(x)\,\mathrm{d}x.$$

The integral on the right-hand side is the area under the famous bell-shaped curve:



The standard normal distribution is also associated with the **Gaussian measure** γ on $\mathbb{R}$; the measure $\gamma(A)$ of a (Lebesgue measurable) set A is

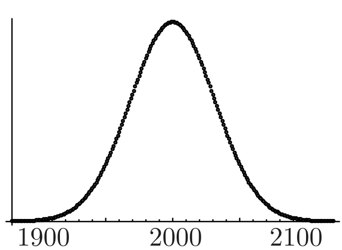
$$\int_A \varphi(x)\,\mathrm{d}x = \mathrm{Prob}[Z \in A],$$

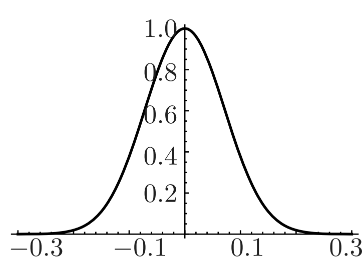
Z again standard normal.

More generally, a random variable X has a **normal distribution** if $X = aZ + b$ for some reals $a \neq 0$ and b, with Z standard normal. Thus, the bell curve can be shifted and re-scaled.

The normal distribution as a limit. The function $e^{-x^2/2}$ may look rather arbitrary (why not $e^{-x^4/4}$, or $\frac{1}{1+x^2}$?), but it is not. The Gaussian distribution appears in mathematics and in sciences incredibly often, usually as a limit or approximation.

Here are two mathematical examples. The left picture below is a plot of the binomial coefficients $\binom{4000}{k}$ for k around 2000, i.e., in the middle:





The right picture shows the function $(\sqrt{1-x^2})^{97}$; this happens to be the density of the mass distribution obtained by projecting the 98-dimensional unit ball onto a line, and also by projecting the uniform measure on the sphere $S^{99} \subset \mathbb{R}^{100}$ onto a line.

In both cases, it can be proved that, with proper formulation, the limiting distribution is normal and so $e^{-x^2/2}$ shows up (if we let 4000 or 97 go to infinity, that is).

Many, although not all, appearances of the normal distribution can be explained by the **central limit theorem**, one of the true wonders of mathematics. Qualitatively speaking it asserts that *if a random variable X is obtained as the sum of many independent influences, each of them relatively small compared to the sum of all others, then X has approximately normal distribution.* There are several quantitative formulations of the central limit theorem, corresponding to various ways of quantifying "relatively small" in this statement, and an interested reader can easily find them in the literature. Since many natural and social phenomena can be regarded as a result of adding up many independent small influences, one encounters approximately normal distributions very often; for example, the errors of measurements are usually normally distributed.

The binomial coefficients from the left picture above are connected to the central limit theorem as follows: Let $X_1, \ldots, X_n$ be independent random variables, each attaining values 0 and 1 with equal probability. Then the central limit theorem implies that the distribution of the sum $S_n = X_1 + \cdots + X_n$, for $n \to \infty$, tends to a normal distribution.[2] At the same time, S_n attains value k with probability $\binom{n}{k}2^{-n}$, since there are $\binom{n}{k}$ ways of selecting the X_i that should attain value 1.

Exercise 3.1. Let $X_1, X_2, \ldots$ be a sequence of independent random variables, each uniformly distributed in $[0,1]$, and let $(u_1, u_2, \ldots)$ be a sequence of real numbers. Look up a statement of the central limit

[2]The technical notion here is *convergence in distribution*. If we set $\tilde{S}_n := (2S_n - n)/\sqrt{n}$, then for every interval $I \subseteq \mathbb{R}$, $\lim_{n\to\infty} \mathrm{Prob}[\tilde{S}_n \in I] = \mathrm{Prob}[Z \in I]$ with Z standard normal.

theorem and derive conditions on the u_i guaranteeing that the distribution of $S_n := \sum_{i=1}^n u_i X_i$ tends to a normal distribution. A geometric interpretation of this result: the projection of the mass of the n-dimensional unit cube to "most" lines tends to the Gaussian measure, up to shift and scaling. ✠

Computing the constant: a double integral trick. We should still compute the constant c in the definition of the density $\varphi(x)$ of the standard normal distribution. In order to get a probability distribution, we must have $\int_{\mathbb{R}} \varphi(x)\,\mathrm{d}x = 1$, and thus $c = 1/I$, where

$$I := \int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x.$$

It is known that the indefinite integral $\int e^{-x^2/2}\,\mathrm{d}x$ is not expressible by elementary functions.[3] Yet I can be evaluated exactly, by, in our opinion, one of the most beautiful computations in all calculus.

The trick is to compute I^2 as

$$\begin{aligned} I^2 &= \left(\int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x\right)\left(\int_{-\infty}^{\infty} e^{-y^2/2}\,\mathrm{d}y\right) \\ &= \int_{\mathbb{R}^2} e^{-x^2/2}e^{-y^2/2}\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{\mathbb{R}^2} e^{-(x^2+y^2)/2}\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_0^{\infty} e^{-r^2/2} 2\pi r\,\mathrm{d}r. \end{aligned}$$

To see the last equality, we consider the contribution of the infinitesimal annulus with inner radius r and outer radius $r + \mathrm{d}r$ to $\int_{\mathbb{R}^2} e^{-(x^2+y^2)/2}\,\mathrm{d}x\,\mathrm{d}y$; the area of the annulus is $2\pi r\,\mathrm{d}r$ and the value of the integrand there is $e^{-r^2/2}$ plus infinitesimal terms which can be neglected.[4] The last integral, $\int_0^\infty e^{-r^2/2} 2\pi r\,\mathrm{d}r$, can already be evaluated in a standard way, by the substitution $t = r^2$, and we arrive

---

[3] There is a precise theorem about this impossibility, and an elegant theory; see, e.g., [**Kas80, MZ94**] for an introduction. This is somewhat paradoxical: while the Gaussian measure is one of the most useful and ubiquitous measures in mathematics, there is no exact formula even for computing the measure of an interval. Working with the Gaussian measure is all about estimates.

[4] This is really an argument for physicists; a completely rigorous one would use a substitution theorem for integration.

at $I^2 = 2\pi$. Thus, the density of the standard normal distribution is $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

Two useful estimates. Sometimes one needs to bound the "tail" of the standard normal distribution, i.e., the function

$$N(z) := \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-x^2/2}\,\mathrm{d}x = \mathrm{Prob}[Z \geq z]$$

with Z standard normal, mainly for large z. Intuitively it seems clear that, since $e^{-x^2/2}$ decreases extremely fast as x grows, the bulk of the integral defining $N(z)$ is contributed by a short interval just after z, and so $N(z)$ should behave more or less like $\varphi(z)$. However, making this rigorous may get messy unless one knows a right trick.

The trick is in finding another function that is "close" to $\varphi(x)$ but can be integrated. Namely, one can use $(1 + \frac{1}{x^2})\varphi(x)$ as an upper bound for $\varphi(x)$, and $(1 - \frac{3}{x^4})\varphi(x)$ as a lower bound. Indeed, these can be integrated by parts, and, e.g., the commercial software Mathematica (or a good calculus student) tells us that

$$\int_x^\infty \left(1 + \frac{1}{x^2}\right) e^{-x^2/2}\,\mathrm{d}x = \frac{e^{-z^2/2}}{z}$$

and similarly for the other function. We obtain

$$\boxed{\left(\frac{1}{z} - \frac{1}{z^3}\right) \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \leq N(z) \leq \frac{1}{z} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad z > 0,}$$

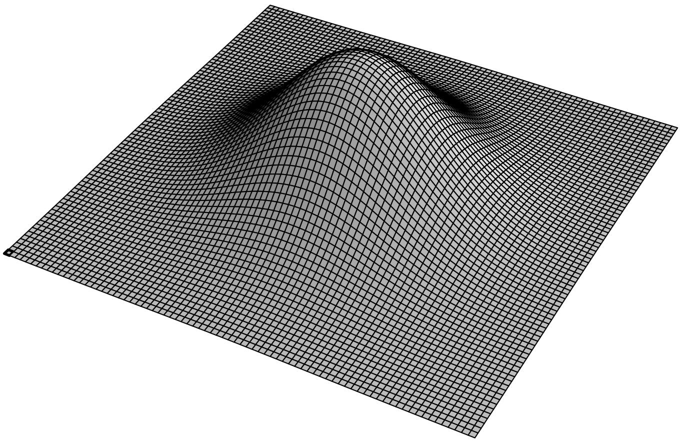
useful formulas to refer to.

The *n*-dimensional Gaussian measure. The Gaussian measure γ on $\mathbb{R}^n$ can be introduced in several ways. It can be obtained as the product of n copies of the 1-dimensional Gaussian measure, and thus the Gaussian measure of a Lebesgue measurable set $A \subseteq \mathbb{R}^n$ is

$$\gamma(A) = (2\pi)^{-n/2} \int_A \prod_{i=1}^n e^{-x_i^2/2}\,\mathrm{d}x = (2\pi)^{-n/2} \int_A e^{-\|x\|^2/2}\,\mathrm{d}x.$$

The latter formula shows that γ is *spherically symmetric*; i.e., its density depends only on the distance to the origin; this was already

used in the computation of the constant c. Here is a plot of the density function in dimension 2:



Probabilistically, the n-dimensional Gaussian measure is obtained as the distribution of a random vector $Z = (Z_1, \ldots, Z_n)$, where $Z_1, \ldots, Z_n$ are independent standard normal random variables.[5]

Indeed, for every product $B = A_1 \times \cdots \times A_n \subseteq \mathbb{R}^n$ of Lebesgue measurable sets, the independence of random variables $Z_1, \ldots, Z_n$ yields

$$\begin{aligned} \operatorname{Prob}[Z \in B] &= \operatorname{Prob}[Z_1 \in A_1 \text{ and } \cdots \text{ and } Z_n \in A_n] \\ &= \prod_{i=1}^{n} \operatorname{Prob}[Z_i \in A_i] = \prod_{i=1}^{n} \gamma(A_i) = \gamma(B), \end{aligned}$$

so the distribution of Z agrees with γ on such products (measurable rectangles), and hence on all measurable sets.

The distribution of Z is also called the ***n*-dimensional standard normal distribution**.

It is hard to overstate the importance of the n-dimensional Gaussian distribution in probability theory and statistics, but also in geometry and algorithms. Here we can give just a small number of

[5]There is a measure-theoretical subtlety here: γ as defined above is a complete measure, but the product of the one-dimensional Gaussian measures is not. To get exactly the same measure, we would need to take a complete product. See Chapter 1, Section 2.3.

simple examples, but as one gets deeper into these fields, applications abound.

Here is a catchphrase underlying many of these applications: *The Gaussian measure in* $\mathbb{R}^n$ *is a better-behaved approximation of the uniform measure on the sphere* S^{n-1}. (Technically speaking, we need to shrink the Gaussian distribution by the factor $\sqrt{n}$ in order to have it match with S^{n-1}.)

This may sound totally insane if one looks at the two-dimensional picture of the Gaussian density above and compares it with the unit circle in the plane. However, this approximation is, again, a high-dimensional phenomenon. In high dimension, the peak of the density of the Gaussian distribution is still at the origin, but the area it occupies is tiny, and as we will see later, most of the mass of γ lies in a thin spherical shell around the sphere of radius $\sqrt{n}$.

A random point on the sphere. Using the Gaussian distribution in $\mathbb{R}^n$, we can easily generate random points on S^{n-1}: we generate $Z = (Z_1, \ldots, Z_n)$, where the Z_i are standard normal and independent, and return the point $Z/\|Z\|$. This assumes, of course, that we can generate independent random numbers with the standard normal distribution, which is not that hard if we have a generator for independent random numbers uniformly distributed in $[0, 1]$.

Projections of the Gaussian distribution and stability. If $Z \in \mathbb{R}^n$ is a random vector distributed according to γ, then, as was mentioned above, the first coordinate Z_1 is standard normal. Since the distribution γ is spherically symmetric, the projection of γ on any line is the 1-dimensional Gaussian measure. Thus, we have the following useful lemma.

Lemma 3.2. *For a random vector* $Z \in \mathbb{R}^n$ *with the* n*-dimensional standard normal distribution and for every fixed unit vector* $u \in S^{n-1}$, *the inner product* $\langle u, Z \rangle$ *has the standard normal distribution.*

Similarly, the orthogonal projection of γ onto a k-dimensional linear subspace of $\mathbb{R}^n$ is the k-dimensional Gaussian measure.

The one-dimensional standard normal distribution is often denoted by $N(0,1)$, and more generally one considers a normal distribution $N(\mu,\sigma^2)$. If Z has the standard normal distribution, which is usually written as $Z \sim N(0,1)$, then $Y = \sigma Z + \mu \sim N(\mu,\sigma^2)$; here μ is the mean (expectation) and σ^2 is the variance, i.e., $\mathbb{E}[(Y-\mu)^2]$.

Exercise 3.3. Check that a standard normal random variable has variance 1. ✠

A fundamental property of the normal distribution, called **stability**, is that the sum of two independent normally distributed random variables is again normally distributed. More precisely, if $X \sim N(\mu,\sigma^2)$ and $Y \sim N(\nu,\tau^2)$, then $X+Y \sim N(\mu+\nu,\sigma^2+\tau^2)$. (In general, a family $\mathcal{D}$ of distributions is called *stable* if every linear combination of independent random variables from the family has again a distribution from $\mathcal{D}$. There are other stable distributions besides the normal one, but the normal distribution is the most important among them.)

A laborious proof of the stability of the normal distribution is by direct computation (we need to compute a convolution). A quick proof follows from the projection property above. Namely, it is easy to see that it is enough to prove that $aX + bY \sim N(0, a^2+b^2)$ for independent $X, Y \sim N(0,1)$; moreover, by re-scaling, we may assume $a^2+b^2 = 1$. The random vector $Z = (X,Y)$ is 2-dimensional standard normal, and so with $u = (a,b)$, we have $\langle u, Z\rangle = aX + bY \sim N(0,1)$ as needed.

The next application of the Gaussian measure is left as an exercise.

Exercise 3.4. (Ball volume via the Gaussian distribution)

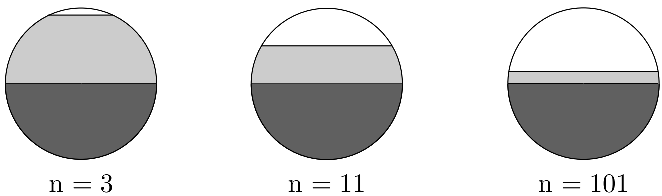
(a) Let $I_n = \int_{\mathbb{R}^n} e^{-\|x\|^2/2}\,\mathrm{d}x$. Calculate I_n using the value of I_1 computed earlier.

(b) Express I_n using $V_n = \lambda(B^n)$, the volume of the n-dimensional unit ball, and a suitable one-dimensional integral, by considering the contribution to I_n of the spherical shell with inner radius r and outer radius $r + \mathrm{d}r$.

(c) Integrating by parts, set up a recurrence and calculate the integral appearing in (b). Compute V_n. ✠

4. Measure Concentration

Let us consider the uniform probability measure μ on the unit sphere S^{n-1} in $\mathbb{R}^n$(see Chapter 1 Section 1.3). We have $\mu(S^{n-1}) = 1$, and the lower hemisphere H (whatever lower means, say $x_n \leq 0$) has measure $\frac{1}{2}$. The following schematic picture shows a neighborhood of H that occupies 90% of the upper hemisphere:



That is, the width of the light gray band is the smallest t such that $\mu(H_t) \geq 0.95$, where, as before, H_t denotes the t-neighborhood of H (measured with the Euclidean distance in $\mathbb{R}^n$).

The point is that as n gets large, a quite small neighborhood of H occupies almost all of the sphere. This is one of the manifestation of measure concentration, but an easy one. The full statement of measure concentration for S^{n-1} says, roughly, that for *every* (measurable) set $A \subseteq S^{n-1}$ of measure $\frac{1}{2}$, a small neighborhood A_t occupies all but an extremely small fraction of the sphere.

There is nothing special about $\frac{1}{2}$ here—for any positive constant the result is similar; see Exercise 4.2 below. For example, if our globe, of radius about 10^4 km, had dimension 10^4 instead of 3, then there would be no way of placing two countries, each occupying over 1% of the surface,[6] more than 700km apart. (On the other hand, land would be real cheap on that globe, since its surface area would be about $1.33 \cdot 10^{26160}$ of, well, 9999-ic kilometers.)

Here is a quantitative statement.

[6]Russia occupies about 3% of our globe's surface, and Canada, United States, China, Brazil, and Australia each between 1.5 and 2%. All other countries have less than 1%.

Theorem 4.1 (Measure concentration on the sphere)**.** *For every $t > 0$ and every $A \subseteq S^{n-1}$ with $\mu(A) \geq \frac{1}{2}$, we have*

$$1 - \mu(A_t) \leq 2e^{-t^2 n/2}.$$

In probabilistic terms, a random point of S^{n-1} avoids the t-neighborhood of A with probability at most $2e^{-t^2 n/2}$.

Thus, for $t = 10/\sqrt{n}$, say, the fraction not covered by A_t is really tiny.

We will discuss proofs of the theorem later.

Exercise 4.2. Derive the following from Theorem 4.1: If $A \subseteq S^{n-1}$ satisfies $\mu(A) \geq \alpha$, $0 < \alpha \leq \frac{1}{2}$, then $1 - \mu(A_t) \leq 2e^{-(t-t_0)^2 n/2}$, where t_0 is such that $2e^{-t_0^2 n/2} < \alpha$, and $t > t_0$. ✠

4.1. More Spaces with Measure Concentration. Whenever we have a space with a notion of distance and with a probability measure, we can ask if there is a measure concentration, that is, whether a small neighborhood of every set of measure $\frac{1}{2}$ occupies all but a tiny fraction of the space.

Measure concentration is quite widespread; intuitively, whenever the considered space has large dimension in some reasonable sense, we can expect measure concentration. The following table summarizes several typical and useful measure concentration results.

Space	Metric	Measure	Concentration
S^{n-1}	Euclidean	μ, uniform	$2e^{-t^2 n/2}$
$\mathbb{R}^n$	Euclidean	γ, Gaussian	$e^{-t^2/2}$
Hamming cube $\{0,1\}^n$	$\lvert\{i : u_i \neq v_i\}\rvert$	$\mu(A) = \frac{\lvert A\rvert}{2^n}$	$2e^{-t^2/2n}$
Permutations of $\{1, \ldots, n\}$	$\lvert\{i : \pi(i) \neq \sigma(i)\}\rvert$	$\mu(A) = \frac{\lvert A\rvert}{n!}$	$e^{-t^2/32n}$
Constant-degree expander G	graph distance	$\mu(A) = \frac{\lvert A\rvert}{\lvert V(G)\rvert}$	e^{-ct}

Each entry of the last column is an upper bound on the maximum possible measure of the complement of A_t in the relevant space for a set A of measure at least $\frac{1}{2}$.

The first row in the table is Theorem 4.1, and the second is self-explanatory. The **Hamming cube** is the discrete space $\{0,1\}^n$, consisting of all strings of 0s and 1s of length n, with the **Hamming metric**, where the distance of two strings is the number of positions where they differ. We can also regard the Hamming cube as a subspace of $\mathbb{R}^n$ with the ℓ_1 norm $\|.\|_1$, where $\|x\|_1 = \sum_{i=1}^n |x_i|$.

Permutations are bijective mappings of the set $\{1,\ldots,n\}$ into itself, and the distance is again the number of i's where they differ.

For the last row of the table, we recall that the *edge expansion* of a graph G with vertex set V is

$$\min\left\{\frac{e(A,V\setminus A)}{|A|} : A\subset V,\, 1\le |A|\le \tfrac{1}{2}|V|\right\},$$

where $e(A,B)$ is the number of edges of G going between A and B. A **constant-degree expander** is a graph G for which the edge expansion is bounded from below by a constant $c_0>0$, maximum degree is bounded above by a constant D, and the number n of vertices is large.[7] In the concentration bound, $c>0$ is a constant depending on c_0 and D but not on n.

Exercise 4.3. If expanders do not look too alien to you, prove the concentration inequality in the table. ✠

4.2. Proofs: the Way of Isoperimetry. Historically, the first measure concentration results were derived according to the following scheme:

(i) (Solve the isoperimetric problem) Among all sets A of a given measure, identify those that minimize the measure of A_t. (In Theorem 2.4, we have proved that for the Euclidean space with the Lebesgue measure these sets are balls, but the measure of the whole space is infinite, so we cannot expect a measure concentration results in the above spirit.)

[7]Strictly speaking, one should consider a *sequence* of graphs, for which c_0 and D fixed while $n\to\infty$.

(ii) (Calculation) For the sets identified in (i), bound the measure of the complement of A_t.

Part (i) of these proofs is usually challenging, while (ii) may be laborious but usually more or less routine.

For S^{n-1}, the isoperimetric problem is solved by **spherical caps**, i.e., balls in S^{n-1} with the Euclidean metric (or expressed differently, intersections of S^{n-1} with halfspaces).

For the Hamming cube, the solutions are also balls, this time in the Hamming metric, provided that there are balls of the given measure (for example, in the 3-dimensional Hamming cube, balls have measures only $\frac{1}{8}$, $\frac{1}{2}$, $\frac{7}{8}$, and 1); if not, then the answer are almost balls but slightly modified.

For the Gaussian measure in $\mathbb{R}^n$, the isoperimetric problem is solved by halfspaces, *not* balls this time.

None of these three results has a proof as simple as the one we have seen for Euclidean isoperimetry. The spherical and Gaussian cases take several pages; the case of the Hamming cube is somewhat simpler.

For more complicated spaces, solving the isoperimetric problem may be too challenging. Fortunately, other methods for proving measure concentration have been developed; as a consequence, one can often get approximate isoperimetric inequalities. We will discuss some of these methods later.

Exercise 4.4. Let $C = C(t)$ be the spherical cap of height $1 - t$:

$$C(t) = \{x \in S^{n-1} : x_1 \geq t\}, \quad -1 \leq t \leq 1.$$

Use the equality $\mu(C) = \lambda(\tilde{C})/\lambda(B^n)$, where $\tilde{C}$ is the cone obtained by connecting all points of C by segments to the origin. Assuming that $t \geq t_0$ for a suitable $t_0 > 0$, enclose $\tilde{C}$ in a ball of radius $\sqrt{1-t^2}$, and derive $\mu(C(t)) \leq e^{-t^2 n/2}$. ✠

Exercise 4.5. A set $N \subseteq S^{n-1}$ is called **ε-dense** if every $x \in S^{n-1}$ is at distance at most ε from some $y \in N$.

(a) Show that any 1-dense set in S^{n-1} has at least $\frac{1}{2}e^{n/8}$ points.

(b) Show that for every $\delta \in (0, 1]$ there is a δ-dense set in S^{n-1} of size at most $(4/\delta)^n$. Hint: consider an inclusion-maximal set in

which every two points have distance greater than δ, and (Euclidean) balls of radius $\delta/2$ around each of its points. ✠

Exercise 4.6. Let P be a convex polytope in $\mathbb{R}^n$ obtained as the intersection of N (closed) halfspaces and containing the unit ball B^n. The goal is to prove that the volume of the ball is very small compared to the volume of P, unless N is exponentially large. Namely,

$$\frac{\lambda(B^n)}{\lambda(P)} \le \left(\frac{C \ln N}{n}\right)^{n/2}$$

for a suitable constant C. Hint: consider a sphere S' of radius $r > 1$ such that the halfspaces complementary to those defining P together cover half of S'. (With a more difficult proof, the factor $\ln N$ can be replaced with $\ln \frac{N}{n}$.) ✠

Exercise 4.7. Continuing the example in Section 1, derive a concentration result for the inner product $\langle u, v \rangle$ of two random vectors in S^{n-1}. ✠

Exercise 4.8. Estimate the probability that a random line through the origin meets the ball of unit radius centered at the point $(2, 0, 0, \ldots, 0)$. It suffices to give a very rough estimate, in order to understand how big or small this quantity may be for n tending to infinity. ✠

Exercise 4.9. Consider two unit balls B, B' in $\mathbb{R}^n$. What is the largest distance of their centers so that a random point in B has a reasonable chance, say at least 1%, of being also in B'? The goal is to obtain a crude answer by simple estimates. ✠

Exercise 4.10. Consider the unit cube $[0,1]^n$, n large, and take the hyperplane h given by the equation $x_1 + \cdots + x_n = \frac{n}{2}$ in the role of an "equator". Estimate, again quite roughly, the width of a parallel strip along h containing 99% of the volume of the cube. Does the answer change if, instead of the volume, we count the vertices? ✠

4.3. Concentration of Lipschitz Functions. We recall the notion of Lipschitz function; this is a strong quantitative version of continuity, which is very important in discrete applications.

A mapping $f\colon X \to Y$ between two metric spaces is called ***C*-Lipschitz** for a real number $C > 0$ if the distance of $f(x)$ and $f(y)$ is never larger than C times the distance of x and y. A mapping is **Lipschitz** if it is C-Lipschitz for some C.

For example, the map $f\colon S^{n-1} \to \mathbb{R}$ assigning to every point its first coordinate is 1-Lipschitz. Another basic example arises by fixing a set A in some metric space X and defining $f(x)$ as the distance of $x \in X$ from A.

One of the most useful manifestations of measure concentration is concentration of Lipschitz functions. Roughly speaking, if X is a space with measure concentration and $f\colon X \to \mathbb{R}$ is a 1-Lipschitz function, then f is almost constant on almost all of X. For example, assuming that the average temperature varies by at most one degree per 100 km on our hypothetical 10^4-dimensional globe of radius 10^4 km, a result below implies that on 99% of the surface the average temperature falls into a 7-degree interval, although the extremes might be as far as 200 degrees apart.

We state the result for the sphere, but exactly the same reasoning works for every space with measure concentration.

Theorem 4.11 (Concentration of Lipschitz functions—Lévy's lemma)**.** *Let $f\colon S^{n-1} \to \mathbb{R}$ be 1-Lipschitz. Then there is $m \in \mathbb{R}$ such that for every $t \in (0,1]$,*

$$\mu\{x \in S^{n-1} : f(x) \notin [m-t, m+t]\} \le 4e^{-t^2 n/2}.$$

Proof. A suitable value of m is the *median* of f, defined as

$$m := \sup\Big\{z : \mu\{x \in S^{n-1} : f(x) \le z\} \le \tfrac{1}{2}\Big\}.$$

We set

$$L := \{x : f(x) \le m\}, \quad H := \{x : f(x) \ge m\}.$$

Soon we will check that both L and H have measure at least $\frac{1}{2}$. Once we know this, we observe that, since f is 1-Lipschitz, we have $f(x) \le m + t$ for all $x \in L_t$, and hence $\mu\{x : f(x) > m+t\} \le 1 - \mu(L_t) \le 2e^{-t^2 n/2}$ by Theorem 4.1. The same argument with H shows $\mu\{x : f(x) < m-t\} \le 2e^{-t^2 n/2}$.

To see that $\mu(H) \geq \frac{1}{2}$, we observe that $S^{n-1} \setminus H$ is a countable union of the sets $A_n := \{x : f(x) \leq m - \frac{1}{n}\}$, each of measure at most $\frac{1}{2}$ (by the definition of m), and so $\mu(S^{n-1} \setminus H) \leq \frac{1}{2}$ too, by countable additivity of the measure μ. The argument for $\mu(L) \geq \frac{1}{2}$ is similar. □

Exercise 4.12. By a simple argument, derive Theorem 4.1 (measure concentration), perhaps with a worse constant in front of $e^{-t^2 n/2}$, from Theorem 4.11 (concentration of Lipschitz functions). ✠

A disadvantage of Lévy's lemma as stated above is that it does not provide the location of the short interval in which the values of f are concentrated. From the proof we get that the concentration is around the median of f, but the median is usually not easy to compute. The next exercise shows that it is enough to compute the average (expectation) of f.

Exercise 4.13. Using Lévy's lemma, prove that if $f\colon S^{n-1} \to \mathbb{R}$ is 1-Lipschitz, then the average $\int_{S^{n-1}} f(x)\,\mathrm{d}\mu(x)$ (or, the expected value of f at a random point of S^{n-1}) is at most $O(n^{-1/2})$ away from the median of f as defined in the proof of Theorem 4.11 above. ✠

Gromov's sphere waist theorem. The Lipschitz condition is crucial for the concentration result.

Exercise 4.14. Construct a continuous function $f\colon S^n \to [0,1]$ that is not concentrated at all. More precisely, for every n and every $x \in [0,1]$, the set $f^{-1}(x - \frac{1}{3}, x + \frac{1}{3})$ has measure at most $\frac{2}{3}$. Hint: start with a very simple function attaining only values 0 and 1. ✠

But, interestingly, there is a deep concentration-style result applying to all continuous functions $S^n \to \mathbb{R}^k$, although the conclusion is different and weaker than in Lévy's lemma. It was essentially proved by Gromov, and a corrected and detailed proof can be found in [**Mem11**].

Theorem 4.15 (Gromov's sphere waist theorem)**.** *Let $f\colon S^n \to \mathbb{R}^k$, $k \leq n$, be continuous. Then there exists $z \in \mathbb{R}^k$ such that for every $\varepsilon > 0$ the ε-neighborhood of $f^{-1}(z)$ in S^n (under the Euclidean metric) has "large" measure, namely, at least $\mu((S^{n-k})_\varepsilon)$, where $(S^{n-k})_\varepsilon$ is the ε-neighborhood of a $S^{n-k} \subseteq S^n$.*

Note that both S^n and S^{n-k} have unit radius, so the S^{n-k} has to be "equatorial."

To compare the result with Lévy's lemma, we look at the case $k = 1$. Lévy's lemma tells us that for a Lipschitz $f\colon S^n \to \mathbb{R}$, there is a $z \in \mathbb{R}$ such that the f-preimage of an ε-neighborhood of z, ε small, occupies nearly all of S^n. So we take a neighborhood in $\mathbb{R}$ and then the preimage. The waist theorem implies that for a continuous f, there is a $z \in \mathbb{R}$ such that the ε-neighborhood of $f^{-1}(z)$, ε small, occupies nearly all of S^n. We take the preimage first and then the neighborhood.

4.4. Gaussian Concentration: the Prékopa–Leindler Way. We present a very elegant proof of a weaker version of the Gaussian concentration of measure, with $2e^{-t^2/4}$ instead of $e^{-t^2/2}$, which is fully sufficient for most applications.

Theorem 4.16 (Approximate Gaussian concentration)**.** *For every $t > 0$ and every measurable $A \subseteq \mathbb{R}^n$ with $\gamma(A) = \frac{1}{2}$, we have $1 - \gamma(A_t) \le 2e^{-t^2/4}$.*

Proof. We will actually prove the following, for every measurable $A \subseteq \mathbb{R}^n$:

$$\int_{\mathbb{R}^n} e^{\operatorname{dist}(x,A)^2/4}\, \mathrm{d}\gamma(x) \le \frac{1}{\gamma(A)}, \tag{3}$$

where $\operatorname{dist}(.,.)$ is the Euclidean distance.

Why would one want anything like this inequality? We need to prove the complement of A_t small. For $x \notin A_t$ we have $\operatorname{dist}(x, A) > t$, hence the integrand in (3) is very large, and so the integral cannot be small unless $\gamma(\mathbb{R}^n \setminus A_t)$ is very small.

Quantitatively, assuming $\gamma(A) = \frac{1}{2}$, we have

$$e^{t^2/4}\gamma(\mathbb{R}^n \setminus A_t) \le \int_{\mathbb{R}^n} e^{\operatorname{dist}(x,A)^2/4}\, \mathrm{d}\gamma(x) \le \frac{1}{\gamma(A)} = 2,$$

and so $\gamma(\mathbb{R}^n \setminus A_t) \le 2e^{-t^2/4}$ as desired.

It remains to prove (3), and for this we use the Prékopa–Leindler inequality with the following magical choices (where χ_A denotes the

characteristic function of A, with $\chi_A(x) = 1$ for $x \in A$ and $\chi_A(x) = 0$ for $x \notin A$):

$$\begin{aligned} f(x) &:= e^{\mathrm{dist}(x,A)^2/4}\varphi(x), \\ g(x) &:= \chi_A(x)\varphi(x), \\ h(x) &:= \varphi(x), \end{aligned}$$

and $t = \frac{1}{2}$ (this t is the exponent in Prékopa–Leindler and has nothing to do with the t in the Gaussian concentration we want to prove); here $\varphi(x) = (2\pi)^{-n/2}e^{-\|x\|^2/2}$ is the density of γ.

Prékopa–Leindler tells us that (all integrals over $\mathbb{R}^n$)

$$\begin{aligned} 1 &= \gamma(\mathbb{R}^n) = \int h \geq \left(\int f\right)^{1/2}\left(\int g\right)^{1/2} \\ &= \left(\int e^{\mathrm{dist}(x,A)^2/2}\,\mathrm{d}\gamma(x)\right)^{1/2}\gamma(A)^{1/2}, \end{aligned}$$

so we indeed get (3), but we must still work to prove the assumption of the inequality, namely, $h(\frac{x+y}{2}) \geq f(x)^{1/2}g(y)^{1/2}$.

If $y \notin A$, then $g(y) = 0$ and we are fine, so let $y \in A$. Then we estimate $\mathrm{dist}(x, A) \leq \|x - y\|$ and calculate

$$\begin{aligned} f(x)g(y) &= e^{\mathrm{dist}(x,A)^2/4}\varphi(x)\varphi(y) \\ &\leq (2\pi)^{-n}e^{\|x-y\|^2/4-\|x\|^2/2-\|y\|^2/2} \\ &= (2\pi)^{-n}e^{-\|x+y\|^2/4} \\ &= h(\tfrac{x+y}{2})^2, \end{aligned}$$

where the penultimate equality uses the parallelogram rule for the Euclidean norm: $\|x+y\|^2+\|x-y\|^2 = 2(\|x\|^2+\|y\|^2)$. This concludes the proof. □

Exercise 4.17. (Approximate measure concentration on S^{n-1}) We recall that the uniform probability measure μ on S^{n-1} can be defined by $\mu(A) = \lambda(\tilde{A})/\lambda(B^n)$, where $\tilde{A}$ is obtained from A by connecting all points by segments to the origin.

(a) Prove that if $A, B \subseteq S^{n-1}$ have Euclidean distance at least t, then the Minkowski sum $\frac{1}{2}(\tilde{A} + \tilde{B})$ is contained in the ball of radius $1 - t^2/8$ around the origin.

(b) Using (a) and the Brunn–Minkowski inequality in the form (2) with $t = \frac{1}{2}$, prove that $\mu(A_t) \geq 1 - 2e^{-t^2/4}$ for every $A \subseteq S^{n-1}$ with $\mu(A) = \frac{1}{2}$. ✠

4.5. The Johnson–Lindenstrauss Flattening Lemma. First we would like to substantiate a statement made earlier in this chapter, namely, that most of the n-dimensional Gaussian measure γ lies in a thin shell around the sphere of radius $\sqrt{n}$ centered at 0; or in probabilistic terms, that the norm of a random Gaussian vector in $\mathbb{R}^n$ is close to $\sqrt{n}$ with probability close to 1.

A straightforward approach is to integrate the contribution of an infinitesimal shell with inner radius x and outer radius $x + \mathrm{d}x$, as we have already done when computing the normalizing constant for the density of the standard normal distribution, or when computing the ball volume in Exercise 3.4. However, this leads to the integral $\int_0^r x^{n-1} e^{-x^2/2}\, \mathrm{d}x$, which does not come out very nicely, although for even n it can be expressed by elementary functions; for example, according to Mathematica,

```
Integrate[x^11 Exp[-x^2/2], {x, 0, r}]
```

$$3840 - e^{-\frac{r^2}{2}} \left(r^{10} + 10r^8 + 80r^6 + 480r^4 + 1920r^2 + 3840\right).$$

So we will not pursue this approach.

On the other hand, using the Gaussian concentration, we obtain a concentration result at once. Indeed, the function $x \mapsto \|x\|$ is (obviously) 1-Lipschitz, and so by the analog of Lévy's lemma (Theorem 4.11) for the Gaussian measure, there is a number m such that

$$\gamma\{x : \|x\| \notin [m - t, m + t]\} \leq 2e^{-t^2/2}$$

(we are using the strong form of Gaussian concentration with $e^{-t^2/2}$, as mentioned in Section 4.1, but Theorem 4.16 would do as well). So in particular, more than 99% of the mass of γ lies in a shell of thickness 7.

This reasoning does not immediately provide any information about the radius m, which is actually typical for concentration arguments. With appropriate general tools, it is often easy to show that something is concentrated in a short interval, but locating that interval may still be hard.

In our case, we can estimate m by considering the expectation of the squared norm. Namely, for a random Gaussian vector $Z = (Z_1, \ldots, Z_n)$, the coordinates Z_i are standard normal, and hence $\mathbb{E}[\|Z\|^2] = n\,\mathbb{E}[Z_1^2] = n$, because a standard normal random variable has unit variance (Exercise 3.3).

Exercise 4.13 asserted that the expectation of a sufficiently concentrated function is close to its median. A very similar argument also shows that $\sqrt{\mathbb{E}[\|Z\|^2]}$ is close to m, the median of $\|Z\|$. The calculation can be done in many ways, the cleanest one being probably integration by parts. But since we want neither to proceed in a non-rigorous way nor to go into substantiating some analytic steps, we replace integration by summation. In order to show that $\mathbb{E}[\|Z\|^2]$ is not much larger than m^2, we estimate the expectation of the deviation $\mathbb{E}[\|Z\|^2 - m^2]$:

$$
\begin{aligned}
\mathbb{E}[\|Z\|^2 - m^2] &\le \sum_{k=0}^{\infty}\Big((m{+}k{+}1)^2 - m^2\Big) \\
&\qquad\quad \cdot \mathrm{Prob}[m{+}k < \|Z\| \le m{+}k{+}1] \\
&\le \sum_{k=0}^{\infty}(2m(k+1) + (k+1)^2)\,\mathrm{Prob}[\|Z\| > m + k] \\
&\le \sum_{k=0}^{\infty}(2m(k+1) + (k+1)^2)\cdot 2e^{-k^2/2} \\
&= O(m+1),
\end{aligned}
$$

since the sums $\sum_{k=0}^{\infty}(k{+}1)^2 e^{-k^2/2}$ and $\sum_{k=0}^{\infty}(k{+}1)e^{-k^2/2}$ are clearly bounded by some constants. Thus, $n = \mathbb{E}[\|Z\|^2] \le m^2 + C(m+1)$, and hence $m \ge \sqrt{n} - O(1)$. A similar calculation with opposite inequalities yields $m \le \sqrt{n} + O(1)$.

This way of estimating m is conceptually straightforward but somewhat cumbersome in details. A different and quite elegant way will be shown in Section 4.6. We should also remark that the distribution of $\|Z\|$ for an n-dimensional random Gaussian Z is known as the χ-*distribution*, and everything one may need to know about it can be found in the statistics literature—but here we mainly want to demonstrate general methods.

We summarize our findings in the following convenient way.

Lemma 4.18. *There is a constant $c > 0$ such that, if Z is n-dimensional standard normal, then for all $t > 0$,*

$$\mathrm{Prob}[Z \notin [\sqrt{n} - t, \sqrt{n} + t]] \le 2e^{-ct^2}.$$

Indeed, for $t \le 1/c$ the bound holds trivially, and for $t > 1/c$ we can use $t - 1/2c \ge t/2$.

The Johnson–Lindenstrauss lemma. The vertex set of a regular simplex, i.e., $n+1$ points with pairwise distances 1, cannot be realized in a Euclidean space of dimension smaller than n (this can be shown by a neat linear algebra argument using the so-called *Gram matrix*). It may thus come as a surprise that if we want all pairwise distances not exactly 1 but, say, between 0.99 and 1.01, then such a point set can be found in dimension only $O(\log n)$.

The Johnson–Lindenstrauss (flattening) lemma asserts that *every* n-point set in a Euclidean space can be put into space of dimension only $O(\log n)$, if we do not mind a slight distortion of the distances, within a factor of $1 \pm \varepsilon$.

Theorem 4.19 (Johnson–Lindenstrauss lemma)**.** *Let $\varepsilon \in (0, 1)$ be a real number, and let $P = \{p_1, p_2, \ldots, p_n\}$ be a set of n points in $\mathbb{R}^n$. Let k be an integer with $k \ge \frac{C \log n}{\varepsilon^2}$, where C is a sufficiently large absolute constant. Then there are points $q_1, \ldots, q_n \in \mathbb{R}^k$, which can be computed by an efficient randomized algorithm, such that*

$$(1 - \varepsilon)\|p_i - p_j\| \le \|q_i - q_j\| \le (1 + \varepsilon)\|p_i - p_j\|$$

for all $i, j = 1, 2, \ldots, n$.

The best known value of C is 4. There are examples showing that the bound on k cannot be improved below $\frac{c \log n}{\varepsilon^2 \log(1/\varepsilon)}$.

This tells us that whenever we have a problem involving Euclidean distances of some high-dimensional data points, and if we can afford, say, 1% relative error in the distances, then we can work in dimension only logarithmic in the number of points.

For example, how can we represent Euclidean distances of data points $p_1, \ldots, p_n$ in $\mathbb{R}^n$? We could store the point coordinates, which are about n^2 numbers. We could also store all the distances, again

quadratically many. However, by the Johnson–Lindenstrauss lemma, if we are content with an approximate representation, it suffices to store only the $O(n \log n)$ coordinates of the q_i.

Remarkably, only the Euclidean spaces are known to enjoy this "flattening" property. For example, if we take the ℓ_1-norm $\|x\|_1 = \sum_{i=1}^n |x_i|$ instead of the Euclidean one, then there exist a constant $\delta > 0$ and n-point sets in $\mathbb{R}^n$ whose ℓ_1-distances cannot be approximately realized, within a factor of 2, in dimension lower than n^δ.

It is hard to overstate the importance of the Johnson–Lindenstrauss lemma, and mainly the main idea of its proof, in contemporary computer science, in areas such as machine learning, data analysis, information retrieval, approximation algorithms, streaming algorithms, etc. There is a nice book by Vempala [**Vem04**] devoted wholly to this subject, as well as several surveys.

All known proofs of the Johnson–Lindenstrauss lemma use the following strategy: define a suitable kind of *random projection*

$$T\colon \mathbb{R}^n \to \mathbb{R}^k,$$

and show that setting $q_i := T(p_i)$ will work with high probability.

The simplest known proofs are obtained if the random projection T has independent Gaussian coefficients. That is, we choose a random $k \times n$ matrix A, whose entries a_{ij} are independent standard normal, and we define $T\colon \mathbb{R}^n \to \mathbb{R}^k$ by

$$T(x) := \frac{1}{\sqrt{k}} Ax;$$

here $1/\sqrt{k}$ is an appropriate scaling factor.

The Johnson–Lindenstrauss lemma is an immediate consequence of the following result.

Lemma 4.20 (Gaussian random projection lemma)**.** *Let n, k be integers, let $T\colon \mathbb{R}^n \to \mathbb{R}^k$ be the random linear map as above, and let $\varepsilon \in (0,1)$. Then for every $x \in \mathbb{R}^n$, we have*

$$\mathrm{Prob}\Big[(1-\varepsilon)\|x\| \le \|T(x)\| \le (1+\varepsilon)\|x\|\Big] \ge 1 - 2e^{-c\varepsilon^2 k},$$

where c is a suitable positive constant.

The probability in the lemma is with respect to the random choice of T, and there is nothing random about x.

Proof of Theorem 4.19 from Lemma 4.20. The value of k in the Johnson–Lindenstrauss lemma is chosen so that Lemma 4.20 yields $\mathrm{Prob}[(1-\varepsilon)\|x\| \le \|T(x)\| \le (1+\varepsilon)\|x\|] \ge 1 - n^{-2}$ for every fixed x.

Setting $x = p_i - p_j$, we have $T(p_i) - T(p_j) = T(p_i - p_j)$ since T is a linear mapping, and hence $\|T(p_i) - T(p_j)\|$ falls outside $(1 \pm \varepsilon)\|p_i - p_j\|$ with probability at most $1/n^2$, for every choice of i, j. Since there are $\binom{n}{2} < n^2/2$ pairs of points to consider, the union bound shows that setting $q_i = T(p_i)$ works with probability at least $\frac{1}{2}$. In particular, suitable points q_i exist. □

Proof of Lemma 4.20. We may assume that $\|x\| = 1$. Then $Z_i := \sum_{j=1}^{n} a_{ij} x_j$ has the standard normal distribution, by the projection property of the n-dimensional Gaussian distribution (Lemma 3.2).

Hence $Z = (Z_1, \ldots, Z_k)$ has the k-dimensional Gaussian distribution, and by the concentration in the spherical shell (Lemma 4.18), $\|Z\|$ falls in the interval $[\sqrt{k} - t, \sqrt{k} + t]$ with probability at least $1 - 2e^{-ct^2}$. Since $T(x) = Z/\sqrt{k}$, the lemma follows by setting $t = \varepsilon\sqrt{k}$. □

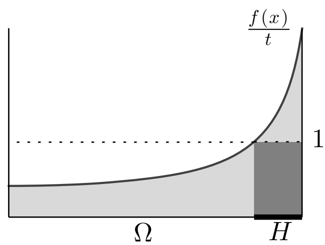
Much effort has been devoted to alternative versions of the random projection lemma, where the random matrix A has various desirable properties. In particular, the main goals are that the matrix can be stored in space much smaller than nk and the random projection T can be evaluated quickly (this is often called a *fast Johnson–Lindenstrauss transform*). One of the first improvements over taking the entries of A independent standard normal was a proof that one can also take independent ± 1 entries, which are faster to generate and calculate with. For a summary of recent work on fast Johnson–Lindenstrauss transforms we refer to Kane and Nelson [**KN14**].

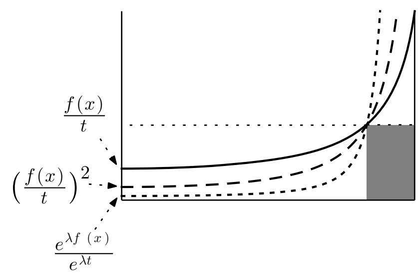
4.6. Product Spaces: the Way of Bernstein and Laplace. We begin with a very useful method of Bernstein from probability theory. Often we need to bound from above the probability that a random variable f attains values larger than some threshold t. In other words,

we have a set Ω with a probability measure μ, and we want to estimate $\mu(H)$, where $H = \{x \in \Omega : f(x) \geq t\}$. Let us assume $t > 0$.

Expressed still differently, we would like to integrate the characteristic function χ_H, but this may be difficult, and so we can try to bound χ_H from above by some other function that is perhaps easier to integrate.

A simple way is to use the function $f(x)/t$, as is indicated in the left picture:





Assuming $f \geq 0$ everywhere, we have $\chi_H(x) \leq f(x)/t$ for all x, and so

$$\operatorname{Prob}[f \geq t] = \int_\Omega \chi_H(x)\,\mathrm{d}\mu(x) \leq \frac{1}{t}\int_\Omega f(x)\,\mathrm{d}\mu(x) = \frac{\mathbb{E}[f]}{t};$$

the result is known as *Markov's inequality*.

In order to get a stronger bound, we can try to make the approximating function steeper near the boundary of H. For instance, we can use $(f(x)/t)^2$, as is indicated in the right picture—this leads to *Chebyshev's inequality* of probability theory.

For a large number of applications, the most successful choice is the function $e^{\lambda f(x)}/e^{\lambda t}$, with a parameter $\lambda > 0$, which is then optimized to get a good bound. The resulting estimate is

$$\operatorname{Prob}[f \geq t] \leq \inf_{\lambda>0} \frac{1}{e^{\lambda t}}\int_\Omega e^{\lambda f(x)}\,\mathrm{d}\mu(x) = \inf_{\lambda>0}\frac{\mathbb{E}[e^{\lambda f}]}{e^{\lambda t}}. \tag{4}$$

This method is usually attributed to Bernstein (who was the first to prove results known as Chernoff's inequality, Hoeffding's inequality, and Azuma's inequality).

Why is this such a good choice of approximating function? First, it is sufficiently nice so that we have a chance to compute or estimate the integral and, in particular, it fits nicely with the Gaussian density and similar functions. Second, it grows fast enough to be almost zero on the region where f is considerably smaller than t, but not too fast: if we had a too fast-growing function, then its integral over H might be too large and kill the whole approach.

Of course, what is too fast depends very much on the specific probability space and f. Using the Bernstein function $e^{\lambda f(x)}/e^{\lambda t}$ typically leads to bounds of the form $\mathrm{Prob}[f \geq t] \leq e^{-at^2}$, with some parameter a. Random variables obeying this kind of bounds are called *subgaussian* (since a standard normal variable is a primary example). Since the subgaussian behavior occurs very often, the Bernstein approach has been very successful.

We actually saw an example of a very similar approach in the proof of the approximate Gaussian concentration from the Prékopa–Leindler inequality in Section 4.4.

A warm-up: sum of independent ±1s. This is probably the simplest example of using the method, and also very well-known. The random variable is $X = X_1 + \cdots + X_n$, the sum of n independent random variables, each attaining values ± 1 with equal probability.

We have

$$\mathbb{E}[e^{\lambda X}] = \mathbb{E}\Big[\prod_{i=1}^{n} e^{\lambda X_i}\Big] = \prod_{i=1}^{n} \mathbb{E}\big[e^{\lambda X_i}\big] = \left(\frac{e^{-\lambda} + e^{\lambda}}{2}\right)^n,$$

where the second equality holds because of independence (or if you prefer, one can write the expectation as an average over all ± 1 vectors, and see directly how it factors into a product).

To proceed further, we need to know, with foresight, that useful values of λ will be much smaller than 1. Then it makes sense to look at the Taylor expansion of $(e^{-\lambda} + e^{\lambda})/2 = \sum_{i=0}^{\infty} \frac{\lambda^{2i}}{(2i)!}$. Seeking an upper bound for this of the form $e^{\text{something}}$, we naturally arrive at $e^{\lambda^2/2}$, for which the Taylor series is the same up to λ^3 and the rest is termwise larger.

Hence

$$\frac{e^{-\lambda}+e^{\lambda}}{2} \le e^{\lambda^2/2}, \tag{5}$$

thus $\mathbb{E}[e^{\lambda X}] \le e^{n\lambda^2/2}$, and so by (4), $\mathrm{Prob}[X \ge t] \le \inf_\lambda e^{n\lambda^2/2-\lambda t}$. Setting $\lambda = t/n$ makes the exponent the most negative, and the resulting bound is the neat $e^{-t^2/2}$. This is usually called *Chernoff's inequality* (or actually, this is its important special case). It is amazing how this simple proof yields a very good bound for a sum of binomial coefficients (since this is what $\mathrm{Prob}[X \ge t]$ really is), which would be very hard to get using the usual estimates of binomial coefficients.

Gaussian measure in a spherical shell revisited. Next, we re-derive a version of Lemma 4.18. For an n-dimensional standard normal vector Z we want to bound $\mathrm{Prob}[\|Z\| \ge \sqrt{n}+t]$, but it will be easier to work with the squared norm, so we will look at $\mathrm{Prob}[\|Z\|^2 \ge Q]$ with $Q = (\sqrt{n}+t)^2$.

To follow the Bernstein recipe we should thus calculate

$$\begin{aligned}\mathbb{E}[e^{\lambda\|Z\|^2}] &= \int_{\mathbb{R}^n} e^{\lambda\|x\|^2}\,\mathrm{d}\gamma(x) = (2\pi)^{-n/2}\int_{\mathbb{R}^n} e^{-(1-2\lambda)\|x\|^2/2}\,\mathrm{d}x\\ &= (1-2\lambda)^{-n/2}\end{aligned}$$

(the substitution $y = \sqrt{1-2\lambda}\,x$ converts the last integral to a multiple of the known integral $\int_{\mathbb{R}^n}\mathrm{d}\gamma(y) = 1$). Note that this works only for $\lambda < \frac{1}{2}$. So by (4), $\mathrm{Prob}[\|Z\|^2 \ge Q]$ is at most

$$\inf_{\lambda\in(0,1/2)} (1-2\lambda)^{-n/2}e^{-\lambda Q} = \inf_{\lambda\in(0,1/2)} e^{-(n/2)\ln(1-2\lambda)-\lambda Q}.$$

Luckily, the exponent can be optimized exactly, with $\lambda = \frac{Q-n}{2Q}$, and we get the bound $e^{-(Q-n)/2}(\frac{Q}{n})^{n/2} = e^{-t\sqrt{n}-t^2/2}(1+\frac{t}{\sqrt{n}})^n$. We estimate $1+\frac{t}{\sqrt{n}} \le e^{t/\sqrt{n}}$, and the whole bound miraculously simplifies to $e^{-t^2/2}$.

Exercise 4.21. Estimate $\mathrm{Prob}[\|Z\| \le \sqrt{n}-t]$ by an analogous method. ✠

Measure concentration: the Hamming cube and other product spaces. The Hamming cube C_n, one of the most prominent spaces in discrete mathematics, can be regarded as an ℓ_1-product:

We define a two-point space $B = \{0, 1\}$, in which the two points have distance 1 and measure $\frac{1}{2}$ each. Then $C_n = B^n$, in the sense that the uniform probability measure on C_n is the product of the measures on the n copies of B, and the distance of two points $u, v \in C_n$ is the sum of the distances of the components, $\mathrm{dist}(u, v) = \sum_{i=1}^{n} \mathrm{dist}(u_i, v_i)$.

In general, if M_i is a space with ground set Ω_i, measure μ_i, and metric dist_i, $i = 1, 2, \ldots, n$, then the **ℓ_1-product** $M_1 \times \cdots \times M_n$ has ground set $\Omega = \Omega_1 \times \cdots \times \Omega_n$, the product measure $\mu = \mu_1 \times \cdots \times \mu_n$, and the distance $\mathrm{dist}(u, v) = \sum_{i=1}^{n} \mathrm{dist}_i(u_i, v_i)$, $u, v \in \Omega$.

We prove a result on concentration of Lipschitz functions on such product spaces, assuming that all factors have small diameter. For convenience, we also assume that each M_i is a probability space, i.e., $\mu_i(\Omega_i) = 1$, so that we can use a probabilistic notation like expectation. In particular, we obtain measure concentration on the Hamming cube.

Theorem 4.22. *Let $M = (\Omega, \mathrm{dist}, \mu) = M_1 \times \cdots \times M_n$ be an ℓ_1-product of metric probability spaces, each of diameter at most 1, and let $f\colon \Omega \to \mathbb{R}$ be a 1-Lipschitz function. Then for every $t > 0$,*

$$\mathrm{Prob}[f \geq \mathbb{E}[f] + t] \leq e^{-t^2/2n}.$$

Of course, applying the theorem to $-f$ also yields $\mathrm{Prob}[f \leq \mathbb{E}[f] - t] \leq e^{-t^2/2n}$.

In the literature, the proof is often done with martingales, which is a nice and somewhat sophisticated probabilistic notion. Here we prove it by clever but elementary use of the Bernstein strategy.

In the examples above, we were bounding $\mathbb{E}[e^{\lambda f}]$ for various particular functions f. Here we want a bound for *all* 1-Lipschitz functions, so it would be good to estimate $\mathbb{E}[e^{\lambda f}]$ for all 1-Lipschitz f simultaneously. However, we need some normalizing condition on f, since otherwise, we can make $\mathbb{E}[e^{\lambda f}]$ arbitrarily large, e.g., by adding a large number to f. But this is all what can happen: a reasonable notion is obtained by requiring $\mathbb{E}[f] = 0$.

We thus define, for a metric probability space $M = (\Omega, \mathrm{dist}, \mu)$, the **Laplace functional** of M as the following function

$E_M\colon [0,\infty) \to [0,\infty]$:

$$E_M(\lambda) := \sup\left\{\mathbb{E}[e^{\lambda f}] : f\colon \Omega \to \mathbb{R} \text{ is 1-Lipschitz and } \mathbb{E}[f] = 0\right\}.$$

Theorem 4.22 is an immediate consequence of the next three lemmas.

Lemma 4.23. *Suppose that* $E_M(\lambda) \le e^{a\lambda^2/2}$ *for some* $a > 0$ *and all* $\lambda > 0$*, and let* $f\colon \Omega \to \mathbb{R}$ *be* 1*-Lipschitz with* $\mathbb{E}[f] = 0$*. Then* $\text{Prob}[f \ge t] \le e^{-t^2/2a}$ *for every* $t > 0$*.*

Exercise 4.24. Prove that using (4). ✠

Lemma 4.25. *Let* $M = (\Omega, \text{dist}, \mu)$ *be a metric probability space of diameter at most* 1*. Then* $E_M(\lambda) \le e^{\lambda^2/2}$ *for all* $\lambda > 0$*.*

Proof. Let f be 1-Lipschitz with zero expectation. The first idea might be to use the fact that the values of f must lie in an interval of length 1 containing 0, but this would only give e^{λ}, not $e^{\lambda^2/2}$. We must use the assumption $\mathbb{E}[f] = 0$ more strongly, and the idea is to integrate over all *pairs* of points—a generally useful trick.

First we recall that *Jensen's inequality* asserts that if $g\colon \mathbb{R} \to \mathbb{R}$ is a convex function and h is a function on some probability space, then $g(\mathbb{E}[h]) \le \mathbb{E}[g \circ h]$, or in integral form, $g\big(\int_\Omega h(x)\,\mathrm{d}\mu(x)\big) \le \int_\Omega g(h(x))\,\mathrm{d}\mu(x)$. Thus we have $e^{\mathbb{E}[h]} \le \mathbb{E}[e^h]$, and so if $\mathbb{E}[h] = 0$, then $\mathbb{E}[e^h] \ge 1$.

Exercise 4.26. Prove $e^{\mathbb{E}[h]} \le \mathbb{E}[e^h]$ directly. ✠

In our case, we use this with $h = -\lambda f$ in order to implant the second variable:

$$\begin{aligned}
\mathbb{E}\big[e^{\lambda f}\big] \quad &\le \quad \mathbb{E}\big[e^{-\lambda f}\big]\,\mathbb{E}\big[e^{\lambda f}\big] \\
&\overset{\text{(Fubini)}}{=} \int\!\!\int e^{\lambda(f(x)-f(y))}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \\
&\overset{\text{(symmetry)}}{=} \int\!\!\int \frac{e^{\lambda(f(x)-f(y))} + e^{\lambda(f(y)-f(x))}}{2}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \\
&\le \quad \int\!\!\int e^{\lambda^2(f(x)-f(y))^2/2}\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y),
\end{aligned}$$

where the last inequality uses the estimate $(e^u + e^{-u})/2 \le e^{u^2/2}$, which is (5) derived in the first example of this section. At this moment we

finally use the diameter bound and the Lipschitz condition: $|f(x) - f(y)| \leq 1$ always, and so $\mathbb{E}\big[e^{\lambda f}\big] \leq e^{\lambda^2/2}$ as desired. $\square$

The last step in the proof of Theorem 4.22 is showing that the Laplace functional is submultiplicative.

Lemma 4.27. *Let $M = M_1 \times M_2$ be an ℓ_1-product of metric probability spaces. Then, for all $\lambda > 0$,*

$$E_M(\lambda) \leq E_{M_1}(\lambda) E_{M_2}(\lambda).$$

Proof. Let $\Omega, \Omega_1, \Omega_2$ be the ground sets of M, M_1, M_2, respectively, and let $f\colon \Omega \to \mathbb{R}$ be 1-Lipschitz with $\mathbb{E}[f] = 0$.

We have

$$\mathbb{E}\big[e^{\lambda f}\big] = \int_{\Omega_2} \int_{\Omega_1} e^{\lambda f(x,y)} \,\mathrm{d}\mu_1(x) \,\mathrm{d}\mu_2(y).$$

For fixed y, we would like to use the definition of E_{M_1} to bound the inner integral, but we cannot do so directly since the slice function $x \mapsto f(x,y)$ need not have zero expectation.

To remedy this, we subtract this expectation and then add it back outside the inner integral. Thus, we define

$$g(y) = \mathbb{E}_x[f(x,y)] = \int_{\Omega_1} f(x,y) \,\mathrm{d}\mu_1(x),$$

and write

$$\mathbb{E}\big[e^{\lambda f}\big] \quad = \quad \int_{\Omega_2} e^{\lambda g(y)} \bigg(\int_{\Omega_1} e^{\lambda (f(x,y) - g(y))} \,\mathrm{d}\mu_1(x) \bigg) \,\mathrm{d}\mu_2(y).$$

After this fix, we get that the inner integral is at most $E_{M_1}(\lambda)$ for every y.

We are left with the integral $\int_{\Omega_2} e^{\lambda g(y)} \,\mathrm{d}\mu_2(y)$. Here we have $\mathbb{E}[g] = \mathbb{E}[f] = 0$, and g is also 1-Lipschitz since it is a weighted average of 1-Lipschitz functions (you may want to check this carefully). Hence the integral is at most $E_{M_2}(\lambda)$, and the lemma, as well as Theorem 4.22, are proved. $\square$

Exercise 4.28. Formulate a version of Theorem 4.22 with the assumption that the diameter of M_i is some $D_i > 0$. ✠

Exercise 4.29. (a) Consider the reals as metric probability space with the usual metric and the Gaussian measure γ. Bound its Laplace functional by $2e^{\lambda^2/2}$. Hint: this does not need any trick, unlike Lemma 4.25.

(b) With more sophistication, can you improve the bound to $e^{\lambda^2/2}$?

(c) Why this, together with Lemmas 4.23 and 4.27, does *not* yield Gaussian measure concentration as in Theorem 4.16? ✠

5. Literature

This chapter owes a great deal to the excellent lecture notes of Ball [**Bal97**]. Our treatment is in many respects similar to Chapters 12–14 in Matoušek [**Mat02**], but thoroughly reworked and hopefully reflecting fifteen more years of experience. On-line lecture notes by Barvinok [**Bar05**] and by Klartag [**Kla13**] give a more advanced picture, including some recent developments. Several chapters in an ambitious textbook in preparation by Hopcroft and Kannan [**HK13**] vividly document the significance of high-dimensional geometry for contemporary computer science. A concise survey from the point of view of a computer scientist, with many recent results (but no proofs) is Vempala [**Vem10**]. Everything one may ever want to know about combinatorial optimization, at least up until 2002, can be found in Schrijver's magnificent book [**Sch03**]

For more material on the Brunn–Minkowski inequality and related development we recommend, e.g., Gardner [**Gar02**]. An impressive application of Brunn–Minkowski to a problem of sorting with partial information, due to Kahn and Linial, can be found, e.g., in [**Mat02**, Chap. 12].

A standard book on measure concentration is Ledoux [**Led01**]. There are also several older surveys, written mainly from the point of view of Banach space theory [**Lin92, LM93, Mil98**], where this notion originated. A concise introduction to concentration of Lipschitz functions and discrete isoperimetric inequalities, including some advanced material and combinatorial applications, can be found in Alon and Spencer [**AS00**] (2nd and later editions).

Acknowledgment. We would like to thank Martin Balko, Zuzana Safernová, and Jana Maříková for reading and valuable comments.

Bibliography

[AS00] N. Alon and J. Spencer. *The Probabilistic Method (2nd edition)*. J. Wiley and Sons, New York, NY, 2000. First edition 1993.

[Bal97] K. Ball. An elementary introduction to modern convex geometry. In S. Levi, editor, *Flavors of Geometry (MSRI Publications vol. 31)*, pages 1–58. Cambridge University Press, Cambridge, 1997.

[Bar05] A. Barvinok. Math 710: Measure concentration. On-line lecture notes, `www.math.lsa.umich.edu/~barvinok/total710.pdf`, 2005.

[Gar02] R. J. Gardner. The Brunn–Minkowski inequality. *Bull. Amer. Math. Soc. (N.S.)*, 39(3):355–405, 2002.

[HK13] J. Hopcroft and R. Kannan. *Foundations of Data Science*. 2013. On-line draft; version from Nov. 2013 retrieved at `http://www.cs.cornell.edu/jeh/book112013.pdf`.

[Kas80] T. Kasper. Integration in finite terms: the Liouville theory. *Math. Mag.*, 53(4):195–201, 1980.

[Kla13] B. Klartag. Regularity through convexity in high dimensions (notes taken by S. Brazitikos). Notes to Winter School Inst. H. Poincaré, Paris, January 7–11,2013, online at `http://wiki-math.univ-mlv.fr/gemecod/lib/exe/fetch.php/klartag-notes-by-brazitikos.pdf`, 2013.

[KN14] D. M. Kane and J. Nelson. Sparser Johnson–Lindenstrauss transforms. *J. ACM*, 61(1), 2014. Article 4. Preprint arXiv:1012.1577, 2010.

[Led01] M. Ledoux. *The Concentration of Measure Phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. Amer. Math. Soc., Providence, RI, 2001.

[Lin92] J. Lindenstrauss. Almost spherical sections; their existence and their applications. In *Jahresbericht der DMV, Jubilaeumstag., 100 Jahre DMV, Bremen/Dtschl. 1990*, pages 39–61, 1992.

[LM93] J. Lindenstrauss and V. D. Milman. The local theory of normed spaces and its applications to convexity. In P. M. Gruber and J. M. Wills, editors, *Handbook of Convex Geometry*, pages 1149–1220. North-Holland, Amsterdam, 1993.

[Mat02] J. Matoušek. *Lectures on Discrete Geometry*. Springer, New York, 2002.

[Mem11] Y. Memarian. On Gromov's waist of the sphere theorem. *J. Topol. Anal.*, 3(1):7–36, 2011.

[Mil98] V. D. Milman. Surprising geometric phenomena in high-dimensional convexity theory. In A. Balog et al., editors, *European Congress of Mathematics (ECM), Budapest, Hungary, July 22–26, 1996. Volume II*, pages 73–91. Birkhäuser, Basel, 1998.

[MZ94] E. A. Marchisotto and G.-A. Zakeri. An invitation to integration in finite terms. *College Mathematics Journal*, 25(4):295–308, 1994.

[Sch03] A. Schrijver. *Combinatorial Optimization. Polyhedra and Efficiency. Vol. A, B, C*, volume 24 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 2003.

[Vem04] S. Vempala. *The Random Projection Method*. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 65. American Mathematical Society, Providence, RI, 2004.

[Vem10] S. Vempala. Recent progress and open problems in algorithmic convex geometry. In *Proc. IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*, pages 42–64. 2010.

Chapter 3

Fourier Analysis

Joseph Fourier won a prize of the French Academy in 1807 for solving the *heat equation*, which is one of the most basic partial differential equations to this day. For example, he could explain theoretically an old empirical rule—namely, that French climate cellars should be dug 2–3 meters underground. In such a cellar, the time shift of the temperature is about half a year, so it is warmest in winter and coldest in summer.

The main ingredient of Fourier's method was an expansion of a function $f\colon \mathbb{R} \to \mathbb{R}$ into a series of sines and cosines

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \bigl(a_n \sin 2\pi n x + b_n \cos 2\pi n x\bigr), \tag{6}$$

where the a_n and b_n are suitable coefficients, assuming that f was periodic with period 1, i.e., $f(x+n) = f(x)$ for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$. The expansion (6) is called the *Fourier series* of f.

Fourier series remain important, but the modern approach to Fourier analysis[1] is different and more general.

Namely, one considers an abelian group G, such as $\mathbb{R}$, the reals with addition, or $\mathbb{Z}_n$, the integers modulo n with addition, or $\mathbb{Z}_2^n$,

[1] Another term is *harmonic analysis*. Some authors use it synonymously with Fourier analysis, others prefer Fourier analysis for the original setting of Fourier series and Fourier transform and reserve harmonic analysis for the general setting explained below.

which can be thought of as the Boolean cube. Fourier analysis provides a basis of the space of complex functions on G with very good properties. The *Fourier transform* $\widehat{f}$ of f can be considered as an expression of f in this new basis.

For finite G, the space $\mathbb{C}^G$ of all functions $f\colon G \to \mathbb{C}$ is simply a finite-dimensional complex vector space. Then the Fourier basis is indeed a basis of this vector space in the sense of linear algebra, even an orthonormal basis. Thus, in the first part of the exposition we focus on the case of G finite.

For G infinite, the situation is more subtle: $\mathbb{C}^G$ is regarded as a Hilbert space, which brings technical complications, although the basic intuition remains the same. We outline the general picture later on, skipping some details. We will also make precise how the Fourier series fit in the general setting.

Applications. Fourier analysis remains one of the key tools for engineering. The Fourier series of a periodic signal shows the strength of various *frequencies*; for example, one can decompose the sound of a musical instrument into pure single-frequency tones (one can tell a note played on a violin from the same note on flute—because of different distribution of the higher frequencies). This musical connection is the origin of the alternatively used term, harmonic analysis.

Such frequency analysis of a signal is not restricted to music; it is also used in frequency filters, as well as in compact encoding of sounds or pictures (although modern encodings, such as the digital photography standard JPEG2000, rely on a newer tool, namely *wavelets*).

In order to work with a periodic continuous signal on a computer, one needs to discretize it. Usually samples are taken every $\frac{1}{n}$-th of the period, so one really deals with a function $f\colon \mathbb{Z}_n \to \mathbb{R}$. It turns out that the Fourier transform $\widehat{f}$ of such a function can be regarded as a function $\mathbb{Z}_n \to \mathbb{C}$.

One of the crucial steps for making Fourier analysis of such signals practical was the discovery of a class of algorithms known as the *fast Fourier transform*, or FFT, which can evaluate $\widehat{f}$ in the $\mathbb{Z}_n$ setting very rapidly even for n huge. In a special issue of the journal *Computing in Science & Engineering* the FFT was included among

"the ten algorithms with the greatest influence on the development and practice of science and engineering in the 20th century," and computer science students will know it from basic courses.

While we will touch these topics briefly, our main goal is to provide the reader with firm understanding of the mathematical foundations. We also add theoretical applications that are perhaps not so well-known: how one can use the Fourier transform to multiply polynomials, test properties of Boolean functions, or find arithmetic progressions in sufficiently large sets. The underlying theme is that looking at a single object from two different angles (coordinates in two different bases) can be tremendously useful.

We also stress that Fourier analysis belongs to the mainstream tools in many branches of mathematics, as well as in theoretical science. The applications demonstrated here are just a tiny sample.

1. Characters

In this section we explain how the Fourier basis is constructed.

We assume that the reader is familiar with the definition of an abelian[2] group, a vector space, an inner product[3], and a norm.[4] Among the infinite groups, we will be interested mostly in $\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{C}$ (the integers, the reals, and the complex numbers with addition). One slightly less usual group is essential, namely the **circle group** $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$—the unit complex numbers with *multiplication*. The letter $\mathbb{T}$ stands for *torus*, as $\mathbb{T}^2$ is, indeed, the geometric torus.

---

[2]That is, commutative.

[3]We recall that an *inner product* on a complex vector space V is a mapping assigning a complex number $\langle u, v\rangle$ to every two vectors $u, v \in V$ and satisfying the following axioms for every $u, u', v \in V$ and $a \in \mathbb{C}$: (1) $\langle u + u', v\rangle = \langle u, v\rangle + \langle u', v\rangle$ and $\langle au, v\rangle = a\langle u, v\rangle$ (linearity in the first variable); (2) $\langle v, u\rangle = \overline{\langle u, v\rangle}$ (conjugate symmetry, the bar denotes complex conjugation: $\overline{x + iy} = x - iy$ for real x, y); and (3) $\langle u, u\rangle$ is real and nonnegative, moreover $\langle u, u\rangle = 0$ iff $u = 0$ (positive definiteness). The standard inner product on $\mathbb{C}^n$ is given by $\langle u, v\rangle = \sum_{i=1}^{n} u_i\overline{v_i}$ (on $\mathbb{R}^n$, this is often called the *dot product*), and on any n-dimensional complex space we can define an inner product via an isomorphism with $\mathbb{C}^n$, for example.

[4]For a vector space V, a *norm* is any mapping assigning a nonnegative real number $\|v\|$ to every vector $v \in V$ that satisfies the following axioms for $u, v \in V$ and $a \in \mathbb{C}$: (1) $\|u + v\| \le \|u\| + \|v\|$ (the triangle inequality); (2) $\|au\| = |a|\|u\|$; and (3) $\|u\| = 0$ iff u is the zero vector. An inner product defines a norm by $\|u\| := \sqrt{\langle u, u\rangle}$; for the standard inner product this defines the Euclidean norm.

A group isomorphic to $\mathbb{T}$ is $\mathbb{R}/\mathbb{Z}$, the reals where we identify numbers that differ by an integer (reals modulo 1, if you wish), with addition. The isomorphism $e\colon \mathbb{R}/\mathbb{Z} \to \mathbb{T}$ is defined by

$$e(x) = e^{2\pi i x}.$$

The group $\mathbb{R}/\mathbb{Z}$ is a natural place to study periodic functions of a real variable, the topic of the classical Fourier analysis.

We start our exposition with finite abelian groups, since this allows us to concentrate on the main aspects. In Section 7 we will see how the theory changes when we consider infinite groups. The major technical difficulties arise from the need to replace finite sums with infinite sums or with integrals; the convergence of these then becomes an issue. An analog of the Fourier analysis also exists for noncommutative groups G; see Chapter 4 on the representation theory.

Until further notice, we assume that G is a finite abelian group. Probably the simplest finite groups are cyclic, of the form $\mathbb{Z}_n := \mathbb{Z}/n\mathbb{Z}$, the integers modulo n; we assume that the reader is familiar with calculation in them. Fortunately, we do not need much more: every nontrivial finite abelian group is isomorphic to a direct sum

$$\bigoplus_{i=1}^{k} \mathbb{Z}_{n_i} \tag{7}$$

for an integer $k \geq 1$ and a sequence of integers $n_1, \ldots, n_k \geq 2$.

We start by defining functions that will later be shown to constitute the Fourier basis.

Definition 1.1. *Let $G = (G, +, 0)$ be a finite abelian group. We say that a function $\chi\colon G \to \mathbb{T}$ is a* **character** *of G if it is a group homomorphism; that is, if*

(i) $\chi(0) = 1$ *and*

(ii) $\chi(x + y) = \chi(x)\chi(y)$ *for all $x, y \in G$.*

We denote the set of all characters of G by $\widehat{G}$.

Exercise 1.2. (a) If χ is a character of a group G, then $1/\chi$ (which equals $\bar{\chi}$, the complex conjugate of χ) is also a character of G.

(b) Let G be a finite group, and suppose $\chi\colon G \to \mathbb{C}\setminus\{0\}$ satisfies (i) and (ii) above. Prove that χ is a character. ✠

The multiplication in $\mathbb{T}$ naturally defines an operation on characters by pointwise multiplication:

$$(\chi \cdot \xi)(x) := \chi(x)\xi(x), \qquad x \in G.$$

Theorem 1.3 (Pontryagin dual for finite groups)**.** *The set $\widehat{G}$ forms an abelian group—with pointwise multiplication as the operation and the constant* 1 *function (denoted by 1_G) as the identity element. The group $\widehat{G}$ is called the* **Pontryagin dual** *of G, and the character 1_G is the* trivial character.[5]

Proof. Obviously, 1_G satisfies the conditions for a character. Commutativity and associativity of pointwise multiplication follow from properties of $\mathbb{T}$. Thus all we need to do is to verify that for characters χ, ξ the mappings $1/\chi$ and $\chi\xi$ are also characters. For $1/\chi$ this was done in Exercise 1.2(a), so it remains to prove that $\alpha = \chi\xi$ is a character.

Consider $x, y \in G$ and calculate:

$$\begin{aligned}
\alpha(x+y) &= \chi(x+y)\xi(x+y) && \text{by the definition of } \alpha \\
&= \chi(x)\chi(y)\xi(x)\xi(y) && \text{since } \chi,\ \xi \text{ are characters} \\
&= \chi(x)\xi(x)\chi(y)\xi(y) \\
&= \alpha(x)\alpha(y).
\end{aligned}$$

□

We proceed by showing that characters form a basis of $\mathbb{C}^G$; this is the Fourier basis.

Let f and g be functions $G \to \mathbb{C}$.

- The *expectation* of f is the expected value of $f(x)$ when $x \in G$ is selected uniformly at random. Since we assume G is finite, we can write simply[6]

$$\mathbb{E}[f] = \mathbb{E}_{x\in G}[f(x)] := \frac{1}{|G|}\sum_{x\in G} f(x).$$

[5] Also called the principal character or the unit character.

[6] For infinite groups we will replace sums by integrals; see Section 7.

- We can use expectation to define a slightly unusual *inner product*.

$$\langle f, g\rangle := \mathbb{E}[f\overline{g}] = \frac{1}{|G|}\sum_{x\in G} f(x)\overline{g(x)}.$$

  It is easy to verify that this is indeed an inner product. The difference from the standard inner product is in the $\frac{1}{|G|}$ factor.

- The L^2 *norm* is defined in the usual way from the inner product defined above.

$$\|f\|_2 := \sqrt{\langle f, f\rangle} = \sqrt{\mathbb{E}_x[|f(x)|^2]}.$$

- At times we will also use other L^p-norms. Again, they differ from the usual definition only by a constant factor—so the validity of the triangle inequality follows from the triangle inequality for the usual L^p norm, which is valid for every $p \geq 1$:

$$\|f\|_p := \sqrt[p]{\mathbb{E}_x[|f(x)|^p]}.$$

We will denote by $L^2(G)$ the vector space $\mathbb{C}^G$ with the inner product and the L^2 norm defined above. After an innocent-looking (but extremely useful) lemma, we start proving linear-algebraic properties of the characters.

Lemma 1.4. *Let G be a finite abelian group and χ a nontrivial character of G. Then*

$$\mathbb{E}[\chi] = 0.$$

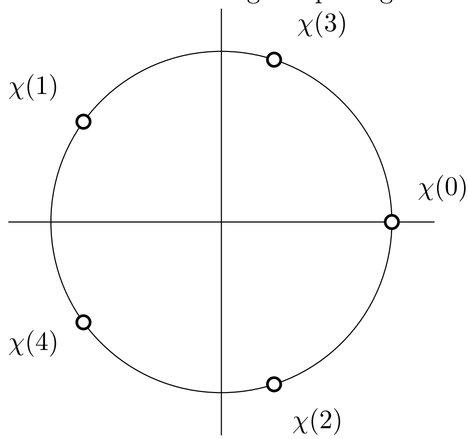
Proof. Fix an arbitrary $y \in G$ and observe that

$$\chi(y)\,\mathbb{E}_{x\in G}[\chi(x)] = \mathbb{E}_{x\in G}[\chi(y)\chi(x)] = \mathbb{E}_{x\in G}[\chi(y+x)] = \mathbb{E}_{x\in G}[\chi(x)].$$

The first equality holds because $\chi(y)$ is a constant. The second one is a part of the definition of a character. Finally, the last equality is true because, as x goes over all elements of G, so does $y + x$.

By rearranging, we get $(\chi(y) - 1)\,\mathbb{E}[\chi] = 0$. It follows that either $\chi(y) = 1$ for every $y \in G$ (which cannot be the case, since χ is nontrivial) or $E[\chi] = 0$. □

While Lemma 1.4 may seem a bit cryptic at first, for a cyclic group it has a natural geometric meaning. In the following picture you can see the values attained by one of the characters of $\mathbb{Z}_5$. Clearly, these values are the vertices of a regular pentagon centered at zero.



Lemma 1.5. *The characters are an orthonormal set of vectors in* $L^2(G)$. *Explicitly:*

(i) $\|\chi\|_2 = 1$ *for every* $\chi \in \widehat{G}$ *and*

(ii) $\langle \chi, \xi \rangle = 0$ *for every* $\chi \neq \xi \in \widehat{G}$.

Proof. The first part is easy: for every $\chi \in \widehat{G}$ and $x \in G$ we have $|\chi(x)| = 1$ and thus also $\mathbb{E}_x |\chi(x)|^2 = 1$. For the second part, let χ and ξ be two distinct characters. Theorem 1.3 implies that $\chi\xi^{-1}$ is also a character; moreover $\chi\xi^{-1} \neq 1_G$. Lemma 1.4 finishes the proof. □

We have proved several properties common to all characters, but so far the only character we have seen is the trivial one. Now we are going to describe all characters of a finite abelian group. We recommend the reader to read the above results (in particular Lemma 1.4) again, checking them for the concrete characters we are about to define.

- For $a, x \in \mathbb{Z}_n$ we define

$$\chi_a(x) := e\left(\frac{a \cdot x}{n}\right) = e^{2\pi i \frac{ax}{n}}.$$

(A trick question: does the dot in $a \cdot x$ represent the product in $\mathbb{Z}_n$ or in $\mathbb{Z}$?) For $a = 0$ we get the trivial character. Moreover, for every $a \in \mathbb{Z}_n$, the mapping χ_a is a character of $\mathbb{Z}_n$. To prove this, observe that $\chi_a(0) = e(0) = 1$ and that

$$\chi_a(x+y) = e\left(\frac{a \cdot (x+y)}{n}\right) = e\left(\frac{a \cdot x + a \cdot y}{n}\right)$$
$$= e\left(\frac{a \cdot x}{n}\right) e\left(\frac{a \cdot y}{n}\right) = \chi_a(x)\chi_a(y).$$

- For $S \subseteq \{1, \ldots, n\}$ and $x \in \mathbb{Z}_2^n$ we let

$$\chi_S(x) := (-1)^{\sum_{i \in S} x_i}.$$

  Every such χ_S is a character of $\mathbb{Z}_2^n$.

- Let G be the group $\bigoplus_{i=1}^k \mathbb{Z}_{n_i}$ for some $k \geq 1$ and some $n_1, \ldots, n_k > 1$. Take $a = (a_1, \ldots, a_k)$, $x = (x_1, \ldots, x_k) \in G$ (that is, $a_i, x_i \in \mathbb{Z}_{n_i}$). We put

$$\chi_a(x) := e\left(\sum_{i=1}^k \frac{a_i x_i}{n_i}\right). \tag{8}$$

  Every mapping χ_a is a character of G; this generalizes the above definitions for $\mathbb{Z}_n$ and for $\mathbb{Z}_2^n$. Moreover, the mapping $a \mapsto \chi_a$ is injective and a group homomorphism (Exercise 1.6).

Perhaps it is good to answer the trick question above explicitly: every element of $\mathbb{Z}_n$ corresponds to some integer, but this integer is determined only up to the addition of a multiple of n. Since we divide by n and use the 1-periodic function $e \colon \mathbb{R} \to \mathbb{T}$, our definition of $\chi_a(x)$ does not depend on the choice of representative. (Otherwise, the definition of χ_a would be very suspicious.) We can use this reasoning for each of a and x separately (and then $a \cdot x$ would be a product of two integers), or for the product $a \cdot x$ (by which we mean the operation in $\mathbb{Z}_n$).

Exercise 1.6. (a) Prove that the mappings χ_S and χ_a given above are characters.

(b) Prove that the assignment $a \mapsto \chi_a$ is an injective group homomorphism.

(c) Observe that the formula for characters of $\mathbb{Z}_2^n$ corresponds to the general formula in Equation (8). ✠

Finally, we are ready to state the properties of the characters as a basis of the vector space $\mathbb{C}^G$. Of course, everybody knows the standard basis of this space (and we will discuss it at the beginning of Section 2). For many applications, though, the Fourier basis is better-behaved, as we will see in the whole of Chapter 3.

Theorem 1.7 (Fourier basis for finite G). *We have*

$$|G| = |\widehat{G}|$$

for any finite abelian group G. Moreover, the mapping $a \mapsto \chi_a$ is a group isomorphism of G and $\widehat{G}$.

The set $\widehat{G}$ is a basis of the vector space $\mathbb{C}^G$.

Proof. Lemma 1.5 implies that elements of $\widehat{G}$ are linearly independent vectors in the $|G|$-dimensional space $\mathbb{C}^G$, and thus $|\widehat{G}| \leq |G|$. The fact that every mapping of the form (8) is a character, and that these characters are distinct, proves the other inequality. This also implies that $\widehat{G}$ is a basis. The easy verification that $a \mapsto \chi_a$ is a group homomorphism is the point of Exercise 1.6(b). □

Next, we present graph-theoretic applications of the results accumulated so far in exercises.

Exercise 1.8. Let G be a finite abelian group and let $S \subseteq G$ be a set such that $0 \notin S$ and S is symmetric (i.e., $S = -S$). The *Cayley graph* $\mathrm{Cay}(G, S)$ is the graph (G, E), where $ab \in E$ whenever $b - a \in S$.

Let χ be a character of G, and let A be the adjacency matrix of $\mathrm{Cay}(G, S)$: $a_{ij} = 1$ if ij is an edge, and $a_{ij} = 0$ otherwise.

(a) Consider a vector $x \in \mathbb{C}^{|G|}$ such that $x_a = \chi(a)$. Prove that x is an eigenvector of $\mathrm{Cay}(G, S)$ (i.e., of the matrix A).

(b) Using (a), find all eigenvalues of the graph C_n, a cycle with n vertices.

(c) Similarly compute all eigenvalues of Q_d, the d-dimensional hypercube: $V(Q_d) = \{0,1\}^d$ and ab is an edge whenever a and b differ in exactly one coordinate. ✠

Exercise 1.9. Check that the graphs from the previous exercise show the (asymptotic) tightness of the following estimates (*Cheeger inequalities*), valid for a d-regular graph $G = (V, E)$.

$$\frac{d - \lambda_2}{2} \leq h(G) \leq \sqrt{2d(d - \lambda_2)}.$$

Here λ_2 is the second largest eigenvalue of G and $h(G)$ is the edge expansion of G defined by

$$h(G) = \min_{\emptyset \neq S \subsetneq V} \frac{|E(S, V \setminus S)|}{\min(|S|, |V \setminus S|)},$$

with $E(A, B)$ standing for the set of all edges connecting A to B. ✠

2. The Fourier Transform

Let us assume that a function $f\colon G \to \mathbb{C}$ is given by a table of values; that is, by the vector $(f(x))_{x \in G}$. In other words, f is given in the standard basis of $\mathbb{C}^G$; i.e., the basis $(\delta_x)_{x \in G}$ with $\delta_x(y) = 1$ for $x = y$ and 0 otherwise. (The function δ_x is called the *Kronecker delta function*.)

The transformation of coordinates from the standard basis to the Fourier basis $\widehat{G}$ is called the Fourier transform, and the backward transformation is known as the inverse Fourier transform.[7] We give two definitions.

Definition 2.1 (Preliminary version). *Let G be a finite abelian group, and let f be a function $G \to \mathbb{C}$. The* Fourier transform *of f is the function $\widehat{f}\colon \widehat{G} \to \mathbb{C}$ defined by*

$$\widehat{f}(\chi) := \langle f, \chi \rangle = \frac{1}{|G|} \sum_{x \in G} f(x) \overline{\chi(x)}.$$

[7] Again, the situation is more subtle for the case of infinite groups.

By Theorem 1.7 the groups G and $\widehat{G}$ are of the same size. A little thought reveals that Equation (8) provides an isomorphism[8] of G and $\widehat{G}$. For this reason it is common to identify $\widehat{G}$ with G using this isomorphism and also consider the Fourier transform of f as a function on G.[9] While Definition 2.1 is conceptually cleaner and also generalizes better to the case of infinite groups, we provide next what is the more usual definition. In the rest of the chapter, we will be using Definition 2.2.

Definition 2.2. *Let G be a finite abelian group, and let f be a function $G \to \mathbb{C}$. The* Fourier transform *of f is the function $\widehat{f}\colon G \to \mathbb{C}$ defined by*

$$\widehat{f}(a) := \langle f, \chi_a \rangle = \frac{1}{|G|} \sum_{x \in G} f(x)\overline{\chi_a(x)}.$$

Exercise 2.3. We assume that an abelian group G and an $a \in G$ are given. We let δ_a be the *Kronecker delta function* defined above, in the first paragraph of this section. Compute the Fourier transform of

(a) χ_a (in particular of χ_0, the constant 1);

(b) δ_a;

(c) $\delta_{-1} + \delta_0 + \delta_1$ for the case $G = \mathbb{Z}_n$; and

(d) $\sum_{i=0}^{k-1} \delta_{i\ell}$ for $G = \mathbb{Z}_{k\ell}$. ✠

Exercise 2.4. Verify that

(a) $\widehat{f}(0) = \mathbb{E}[f]$ for every finite abelian group and every function f.

(b) $\|\widehat{f}\|_\infty \leq \|f\|_1$—explicitly, for every a we have $|\widehat{f}(a)| \leq \mathbb{E}[|f|]$. ✠

[8] However, for instance, the mapping $a \mapsto \chi_{-a}$ is another isomorphism.

[9] In Section 7 we will see that for infinite groups, G and $\widehat{G}$ are not necessarily isomorphic. Then the domain of $\widehat{f}$ will be changed accordingly.

Theorem 2.5 (Inverse Fourier transform)**.** *Let G be a finite abelian group, $f \in \mathbb{C}^G$ any function. Then*

$$f = \sum_{a \in G} \widehat{f}(a)\chi_a.$$

Note that f and χ_a are functions $G \to \mathbb{C}$, while the $\widehat{f}(a)$ are complex coefficients.

Proof. This is a basic linear algebra result, but we prove it for completeness. Put $g = f - \sum_a \widehat{f}(a)\chi_a$. For any character ξ it is easy to verify that $\langle g, \xi \rangle = 0$, since the characters are orthonormal. Because the characters generate the space $\mathbb{C}^G$, it follows that $g = 0$. □

Exercise 2.6. Write the inverse Fourier transform explicitly for the functions from Exercise 2.3. ✠

Now that we know how to go from f to $\widehat{f}$ and back again, we want to study the properties of this bijection.

Theorem 2.7 (Plancherel's)**.** *For every $f, g \in L^2(G)$ we have*

$$\langle f, g \rangle = \sum_{a \in G} \widehat{f}(a)\overline{\widehat{g}(a)}.$$

The proof of Theorem 2.7 is, of course, just a straightforward use of linear algebra; we leave the details for the reader to verify. The result, however, is very useful. In particular, its easy corollary, Parseval's theorem below, is going to be used many times in this chapter.

Note that the right-hand side in Plancherel's theorem is an inner product on $\widehat{G}$—the "usual one," different from the inner product we have defined on G.

Theorem 2.8 (Parseval's)**.** *For every $f \in L^2(G)$ we have*

$$\|f\|_2 = \sqrt{\sum_{a \in G} |\widehat{f}(a)|^2}.$$

As above, we note that the right-hand side of Parseval's theorem defines a norm on $\mathbb{C}^G$—the "usual L^2–norm." To emphasize this

point, we define new inner product and norm on the space $\mathbb{C}^G$:

$$\widehat{\langle f, g\rangle} := \sum_{a\in G} \widehat{f}(a)\overline{\widehat{g}(a)}$$

$$\widehat{\|f\|}_2 := \widehat{\langle f, f\rangle} = \sqrt{\sum_{a\in G} |\widehat{f}(a)|^2}.$$

We also let $\mathcal{F}$ denote the Fourier transform as a mapping: if f is a function on G and $\widehat{f}$ is the Fourier transform of f, we write $\widehat{f} = \mathcal{F}(f)$.

Theorem 2.9. *The Fourier transform $\mathcal{F}\colon f \mapsto \widehat{f}$, considered as a mapping from $L^2(G) = (\mathbb{C}^G, \langle\cdot,\cdot\rangle)$ to $(\mathbb{C}^G, \widehat{\langle\cdot,\cdot\rangle})$ is an isomorphism of vector spaces with inner products. Explicitly, $\mathcal{F}$ is linear, invertible, and preserves inner products.*

Exercise 2.10. Fix $p \in G$ and $c \in \mathbb{C}$. We define operators T_p and P_c to map a function $f\colon G \to \mathbb{C}$ to another function defined as follows:

$$(T_p f)(x) = f(x+p), \qquad (P_c f)(x) = cf(x).$$

In case G is a field and $p \neq 0$, we also define $(S_p f)(x) = f(px)$. Prove that

(a) $\widehat{T_p f}(a) = \chi_a(p)\widehat{f}(a)$;

(b) $\widehat{P_c f}(a) = c\widehat{f}(a)$;

(c) $\widehat{S_p f}(a) = \widehat{f}(a/p)$. ✠

Exercise 2.11. Find the matrix of the linear mapping given by the Fourier transform on $\mathbb{Z}_n$. Explicitly, find a matrix M_n such that for every $f\colon \mathbb{Z}_n \to \mathbb{C}$ we have

$$(\widehat{f}(0), \ldots, \widehat{f}(n-1))^T = M_n(f(0), \ldots, f(n-1))^T.$$

Compute $\det M_n$ and thus re-prove the fact that the Fourier transform is a bijection. ✠

A careful reader may have observed a curious fact in Exercise 2.3. If one of the functions f, $\widehat{f}$ is nonzero only at a single point, then the other is nonzero everywhere. The next exercise generalizes these examples; it may be considered a baby version of the celebrated *uncertainty principle*, originally from quantum mechanics. Various versions of this principle say that f and $\widehat{f}$ can never both be "strongly concentrated."

Exercise 2.12. Let G be an abelian group, and let $f\colon G \to \mathbb{C}$ be a function that is not identically zero. We define the *support* of f to be the set $\mathrm{supp}(f)$ of all $x \in G$ for which $f(x) \neq 0$. Prove that

$$|\,\mathrm{supp}(f)| \cdot |\,\mathrm{supp}(\widehat{f})| \geq |G|.$$

✠

Time complexity. So far we have not addressed the interesting question of how fast we can compute the Fourier transform. For a function f defined on an abelian group with n elements, computing each of the coefficients $\widehat{f}(a)$ using Definition 2.2 clearly requires $O(n)$ arithmetic operations. (We count each operation with a complex number as a unit, avoiding numerical issues or counting bit operations.) This gives an easy way to compute the whole of $\widehat{f}$ in $O(n^2)$ operations. Thus it may come as a surprise that it is possible to do the same in time $O(n \log n)$ using the celebrated *fast Fourier transform* (or FFT for short). This algorithm was known to Gauss in 1805, and it was rediscovered by Cooley and Tukey in 1965. The existence of so fast an algorithm is one of the reasons why the Fourier transform is so useful in practice.

While we will not spend much time on FFT here (many readers probably know it already), we devote the next exercise to FFT for functions on $\mathbb{Z}_2^n$, a case that avoids the technicalities of the more usual case (function on $\mathbb{Z}_N$, in particular with N a power of 2).

Exercise 2.13. Let $f\colon \mathbb{Z}_2^n \to \mathbb{C}$ be a function. Using f we define functions $f_0, f_1\colon \mathbb{Z}_2^{n-1} \to \mathbb{C}$ by

$$f_0(u) = f(u0) \qquad \text{and} \qquad f_1(u) = f(u1),$$

where by $u0$ we mean the concatenation of u and 0, and similarly for $u1$.

(a) For $a \in \mathbb{Z}_2^n$ we put $\bar{a} = (a_1, \ldots, a_{n-1})$. Express $\widehat{f}(a)$ using $\widehat{f_0}(\bar{a})$ and $\widehat{f_1}(\bar{a})$. (You may want to treat the cases $a_n = 0$ and $a_n = 1$ separately.)

(b) Explain how the result of part (a) yields an algorithm that computes all 2^n values of $\widehat{f}(\bar{a})$ in time $O(n2^n)$.

(c) Try the same for functions defined on $\mathbb{Z}_{2^n}$. ✠

3. Two Unexpected Applications

In this section we will meet two applications that require nothing but the basic properties of the Fourier transform. The results are nonetheless rather surprising.

3.1. Linearity Testing. Suppose we have a black box that evaluates a *Boolean function* f of n variables for any given point. We wish to test whether f is linear; that is, whether $f(0) = 0$ and $f(x+y) = f(x) + f(y)$ for every x, y in $\mathbb{Z}_2^n$. Obviously, we want to do this as fast as possible. For simplicity, we measure this by counting how many times we evaluate f.

If we have no prior knowledge about f, then we need to look at $f(x)$ for every $x \in \mathbb{Z}_2^n$ to be certain. Since 2^n is prohibitively high in applications, we may want to try some sort of approximation. An interesting variant was suggested by Blum, Luby, and Rubinfeld [**BLR93**]. They allowed for a small probability of error, and did not care about border-line cases: functions that are close to linear. This approach has led to the development of so-called *property testing*, which is now an important research field.

As a historical remark, the original motivation for Blum et al. was to study in general how to test correctness of a computer implementation of some function. Moreover, a few years later this became an important step in the proof of the celebrated PCP theorem; see, e.g., Arora and Barak [**AB09**] or Goldreich [**Gol08**].

Theorem 3.1. *Let $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$ be a Boolean function and let ε, δ be positive reals. There is a randomized algorithm that evaluates f at $3\lceil \frac{\log \delta}{\log(1-\varepsilon)} \rceil$ points and has the following properties:*

- *If f is linear, the algorithm always confirms it.*
- *If every linear function g differs from f on at least $\varepsilon 2^n$ points, then with probability at least $1-\delta$ the algorithm answers that f is not linear.*

Proof. The algorithm is as simple as it could possibly be: we repeatedly test the linearity. Specifically, we take $N = N(\varepsilon, \delta)$ large. We choose random $x, y \in \mathbb{Z}_2^n$ and test whether $f(x+y) = f(x) + f(y)$;

repeat this N-times, all random choices independent and uniform. If we find no counterexample, we declare f linear.

The algorithm behaves as it should for linear functions. To analyze its behaviour for nonlinear f, let q denote the probability we are fooled at one step; that is, $q = \text{Prob}_{x,y\in\mathbb{Z}_2^n}[f(x+y) = f(x) + f(y)]$. The probability that the algorithm declares f linear equals q^N. We need to find an upper bound on q^N for f that is ε-far from every linear mapping.

To achieve that, we define a mapping $F(x) = (-1)^{f(x)}$. It is easy to see that f is linear if and only if F is a character; the number of points at which f needs to be changed to become linear also coincides with the number of points where F needs to be changed to become a character. For a character χ of $\mathbb{Z}_2^n$ we put $p(\chi) = \text{Prob}_x[F(x) \neq \chi(x)]$. We only care about the behaviour of the algorithm when $p(\chi) \geq \varepsilon$ for every character χ.

We will consider the Fourier transform of F and use several times the basic formula

$$F(x) = \sum_{a\in\mathbb{Z}_2^n} \widehat{F}(a)\chi_a(x). \tag{9}$$

We start by observing that (writing $p = p(\chi_a)$)

$$\widehat{F}(a) = \mathbb{E}_x[F(x)\overline{\chi_a(x)}] = 1\cdot(1-p) + (-1)\cdot p = 1 - 2p \leq 1 - 2\varepsilon.$$

Next, we express q similarly. We will again use the easy fact that two numbers in $\{\pm 1\}$ are equal if their product is 1; otherwise the product is -1. Consequently

$$\mathbb{E}_{x,y}[F(x+y)F(x)F(y)] = 1\cdot q + (-1)\cdot(1-q) = 2q - 1.$$

On the other hand, using (9) we obtain

$$\begin{aligned}
&\mathbb{E}_{x,y}[F(x+y)F(x)F(y)] = \\
&\quad = \mathbb{E}_{x,y}\Big[\sum_{a\in\mathbb{Z}_2^n} \widehat{F}(a)\chi_a(x+y) \sum_{b\in\mathbb{Z}_2^n} \widehat{F}(b)\chi_b(x) \sum_{c\in\mathbb{Z}_2^n} \widehat{F}(c)\chi_c(y)\Big] \\
&\quad = \sum_{a,b,c} \widehat{F}(a)\widehat{F}(b)\widehat{F}(c)\, \mathbb{E}_x[\chi_a(x)\chi_b(x)]\, \mathbb{E}_y[\chi_a(y)\chi_c(y)].
\end{aligned}$$

To simplify further, we recall that the mapping $a \mapsto \chi_a$ is a group homomorphism. Consequently, $\chi_a(x)\chi_b(x) = \chi_{a+b}(x)$. We deal with the second product of characters similarly. Using Lemma 1.4 we find that the only nonzero terms of the last sum are those satisfying $a + b = 0$ and $a + c = 0$. For such values of a, b, and c, the two expectations equal 1, yielding (note that $-a = a$)

$$\mathbb{E}[F(x+y)F(x)F(y)] = \sum_a \widehat{F}(a)^3.$$

Finally, we recall that $\widehat{F}(a) \leq 1 - 2\varepsilon$ and write

$$\mathbb{E}[F(x+y)F(x)F(y)] \leq (1-2\varepsilon)\sum_a \widehat{F}(a)^2 = (1-2\varepsilon)\|F\|_2^2.$$

We are using the fact that $\widehat{F}(a)$ is a real number (thus its square is nonnegative) and, crucially, Parseval's theorem (Theorem 2.8).[10] Since F attains only values ± 1, its norm is 1. To wrap up, we have obtained the inequality $2q-1 \leq 1-2\varepsilon$, and thus $q \leq 1-\varepsilon$. It remains to choose N so that $(1-\varepsilon)^N \leq \delta$, or equivalently, $N \geq \frac{\log \delta}{\log(1-\varepsilon)}$. □

3.2. Arithmetic Progressions. Our next application is from number theory. The history starts in 1952 with Roth's theorem: every sufficiently large set $A \subseteq [N] := \{1, 2, \ldots, N\}$ contains an arithmetic progression of length 3.

What exactly sufficiently large means (i.e., what is the weakest assumption on $|A|$ that will do) is still open. Roth proved that for every $\delta > 0$ and N large (larger than some function of δ), it is enough to assume that $|A| > \delta N$; compare with Exercise 3.3 below. Later, this result was greatly extended by Szemerédi. Using combinatorial methods, he proved that for every k and $\delta > 0$, if N is sufficiently large, $A \subseteq [N]$, and $|A| > \delta N$, then A contains an arithmetic progression of length k. Recently, combining Szemerédi's combinatorial and Roth's Fourier-analytic approach, Green and Tao proved that arbitrarily long arithmetic progressions exist in primes. This development

[10] Observe that it is not enough to use the bound $\widehat{F}(a)^3 \leq (1-2\varepsilon)^3$ for every a, since we get an extra factor of 2^n.

is nicely explained in a survey by Tao [**Tao07**]; in particular similarities of seemingly different approaches are explained there. A more detailed treatment is provided in Tao and Vu [**TV06**].

To explain one of the underlying ideas of this development we present an analogous result for the group $\mathbb{Z}_3^n$ that was proved by Meshulam [**Mes95**] using Roth's ideas. The same proof works for $\mathbb{Z}_p^n$ for any prime p. We restrict to $p = 3$ for the sake of clarity.

Theorem 3.2. *Every set $A \subseteq \mathbb{Z}_3^n$ of size $|A| \geq c\frac{3^n}{n}$ contains an arithmetic progression of length 3 (for a suitable constant $c > 0$).*

Proof. With a slight abuse of notation, let $A(x)$ be the characteristic function of the set A. We will use the Fourier transform of the function A to either find an arithmetic progression of length 3 (an AP_3, for short), or to find a character of $\mathbb{Z}_3^n$ "along which A oscillates a lot." This will allow us to find a hyperplane (i.e., an $(n-1)$-dimensional affine subspace of $\mathbb{Z}_3^n$) in which A has higher density than in the whole of $\mathbb{Z}_3^n$ and induction will finish the proof.

To start slowly, we observe that $\widehat{A}(0) = \mathbb{E}_x[A(x)]$ is the density of the set A; that is, $\widehat{A}(0) = \frac{|A|}{3^n}$. We define $\delta := \frac{|A|}{3^n}$; by assumption, $\delta \geq \frac{c}{n}$.

It will be convenient to consider AP_3's of the form $x-d, x, x+d$, with difference $d \in \mathbb{Z}_3^n$ and starting point $x-d$. The key quantity for the proof is the "density of AP_3's"

$$T(A) = \mathbb{E}_{x,d}[A(x-d)A(x)A(x+d)]. \tag{10}$$

The number of AP_3's is $T(A) \cdot |\mathbb{Z}_3^n|^2$. Of those, many are trivial: the expression for $T(A)$ allows the difference d to be 0, thus it counts 3^n progressions of the form x, x, x. It follows that to prove there is a nontrivial AP_3 we need to show

$$T(A) \cdot 3^{2n} > 3^n.$$

To this end, we substitute the Fourier expansion of A to Equation (10) and use tricks similar to those in the proof of Theorem 3.1.

$$T(A) = \mathbb{E}_{x,d}\Big[\sum_{a\in\mathbb{Z}_3^n}\widehat{A}(a)\chi_a(x-d)\sum_{b\in\mathbb{Z}_3^n}\widehat{A}(b)\chi_b(x)\sum_{c\in\mathbb{Z}_3^n}\widehat{A}(c)\chi_c(x+d)\Big]$$

$$= \sum_{a,b,c\in\mathbb{Z}_3^n}\widehat{A}(a)\widehat{A}(b)\widehat{A}(c)\,\mathbb{E}_x[\chi_a(x)\chi_b(x)\chi_c(x)]\,\mathbb{E}_d[\chi_a(-d)\chi_c(d)]$$

(11)

$$= \sum_{a,b,c\in\mathbb{Z}_3^n}\widehat{A}(a)\widehat{A}(b)\widehat{A}(c)\,\mathbb{E}_x[\chi_{a+b+c}(x)]\,\mathbb{E}_d[\chi_{c-a}(d)]$$

$$= \sum_{a\in\mathbb{Z}_3^n}\widehat{A}(a)^2\widehat{A}(-2a)$$

To estimate this further, we put $M = \max_{b\neq 0}|\widehat{A}(b)|$. We use the triangle inequality, Parseval's theorem, and the easy fact $\|A\|_2^2 = \delta$ to obtain

$$|T(A) - \delta^3| = \Big|\sum_{0\neq a\in\mathbb{Z}_3^n}\widehat{A}(a)^2\widehat{A}(-2a)\Big| \le \sum_{0\neq a\in\mathbb{Z}_3^n}|\widehat{A}(a)|^2\cdot M \le \delta M.$$

We distinguish two cases:

Case 1: $M \le \delta^2/2$. We conclude that $T(A) \ge \delta^3 - \delta M \ge \delta^3/2$. If $3^n\delta^3/2 > 1$, we are done. This requires $3^n c^3/n^3 > 2$, which holds for every $c \ge 2$.

Two related functions can help illustrate this case. Let A_1 be the constant function $A_1(x) \equiv \delta$. Easily, we have $T(A_1) = \delta^3$. More importantly, suppose that for every x we have a random variable $A_2(x)$ such that $A_2(x) = 1$ with probability δ and $A_2(x) = 0$ otherwise. Let all of these 3^n random variables (for all $x \in \mathbb{Z}_3^n$) be independent. Formally, there is some large probability space Ω and for every x a mapping $A_2(x) : \Omega \to \{0, 1\}$. For every x, the expected value of $A_2(x)$ is δ; formally, $\mathbb{E}_{\omega\in\Omega}[A_2(x)(\omega)] = \delta$. Then $T(A_2)$ is a random variable with expectation approximately δ^3 (why not exactly δ^3?). To wrap up, $A_1(x)$ is equal to δ for every x and $A_2(x)$ has expected value δ for every x. The condition on $A(x)$ is much weaker: we only know that the average *over all* x is equal to δ. In Case 1 of the proof we have shown that if all nontrivial Fourier coefficients of A are rather

small, then A still resembles the functions A_1, A_2 defined above—in the sense that $T(A)$ is of the same order as $T(A_1)$ and $T(A_2)$. Such considerations are frequently useful to motivate the direction of a proof.

Case 2: $M \geq \delta^2/2$. Now is the time we use the large Fourier coefficient to find a hyperplane H such that $|A \cap H|/|H| \geq \delta + \delta^2/4$. This will enable us to do the inductive step and finish the proof.

To start filling in the details, let $a \in \mathbb{Z}_3^n$ be such that $|\widehat{A}(a)| \geq \delta^2/2$ and $a \neq 0$. We recall that $\chi_a(x) = e(\sum_i a_i x_i/3)$. We write $a \cdot x = \sum_i a_i x_i$ and $\omega = e(-1/3)$. With this notation, we may write $\widehat{A}(a)$ in a succinct form

$$\begin{aligned}\widehat{A}(a) &= \mathbb{E}_x[A(x)\overline{\chi_a(x)}] \\ &= \mathbb{E}_x[A(x)\omega^{a\cdot x}] \\ &= \mathbb{E}_{t\in\mathbb{Z}_3}\left[\mathbb{E}_{x:a\cdot x=t}[A(x)\omega^t]\right] \\ &= \mathbb{E}_{t\in\mathbb{Z}_3}[\alpha_t\omega^t],\end{aligned}$$

where $\alpha_t := |\{x \in A : a \cdot x = t\}|/3^{n-1}$ is the density of A in a certain hyperplane. To summarize, we have

$$\frac{\alpha_0+\alpha_1+\alpha_2}{3} = \delta \qquad \text{and} \qquad \frac{\left|\alpha_0+\alpha_1\omega+\alpha_2\omega^2\right|}{3} \geq \frac{\delta^2}{2} \tag{12}$$

(the first equality is the density of A expressed in two ways). Roughly speaking: the average of the α_t is δ, and not all of them can be equal, since $\sum_t \omega^t = 0$. To quantify this line of reasoning, let us write $\alpha_t = \delta + \beta_t$. Rewriting the above,

$$\frac{\beta_0+\beta_1+\beta_2}{3} = 0 \qquad \text{and} \qquad \frac{\left|\beta_0+\beta_1\omega+\beta_2\omega^2\right|}{3} \geq \frac{\delta^2}{2}.$$

Using the triangle inequality and adding the two equations we obtain

$$\frac{1}{3}\sum_{t\in\mathbb{Z}_3}(|\beta_t|+\beta_t) \geq \delta^2/2$$

and, consequently, for some t we have $|\beta_t| + \beta_t \geq \delta^2/2$. Thus $\alpha_t \geq \delta + \delta^2/4$, as promised.

Now we get to the big picture. We plan to use induction on n. For $n = 1$ the result is true if $c \cdot 3^1/1 \geq 3$, or $c \geq 1$. For $n > 1$ we use the computations above and conclude that either the result is true

(in Case 1, if $c \geq 2$), or that we can apply the inductive assumption. (Here we use the fact that $\mathbb{Z}_3$ is a field, and thus each hyperplane in $\mathbb{Z}_3^n$ is isomorphic to $\mathbb{Z}_3^{n-1}$.) For this we need that

$$\delta + \delta^2/4 \geq \frac{c}{n-1}.$$

Since $\delta \geq c/n$, having $c \geq 4n/(n-1)$ is sufficient; this is most restrictive for $n = 2$. Putting all conditions together, $c = 8$ in the statement is sufficient. □

Another way to state the proof is perhaps intuitively more convincing: if we do not find the desired progression by direct counting, we find a subspace where the set A is considerably more dense than in the whole space. This process cannot go forever, though, since we would have found a set in which the density of A is larger than 1, which is clearly impossible.

The original Roth's theorem speaks about arithmetic progressions in an initial segment of positive integers. It uses the same main idea: small Fourier coefficients imply many arithmetic progressions, while at least one large coefficient will enable us to find a place where A is denser than in the whole set. There are numerous technical difficulties, though. A nice account of this proof appears in Tao and Vu [**TV06**].

In the next exercise, the reader can rediscover the properties of Behrend's construction that provides a lower bound for possible improvement of Roth's theorem. The best known lower bound is due to Elkin (nicely explained in [**GW10**]): a subset of $[N]$ with $\Omega(N(\log N)^{1/4}/2^{2\sqrt{2\log_2 N}})$ points and no 3-term arithmetic progression. This is still rather far from the best known upper bound in the improvement of Roth's theorem due to Bourgain: for some C, a set of size at least $CN(\log\log N/\log N)^{1/2}$ contains a 3-term arithmetic progression.

Exercise 3.3. (a) Prove that there is a set $A \subseteq [0, d]^n \cap \mathbb{Z}^n$, which contains at least $\frac{(d+1)^n}{nd^2}$ points on a sphere centered at the origin.

(b) Verify that the set of integers $\{\sum_{i=0}^{n-1} x_i(2d+1)^i : (x_i)_{i=0}^{n-1} \in A\}$ contains no 3-term arithmetic progression.

(c) Conclude that there is a set in $\{1, \ldots, N\}$ that contains no 3-term arithmetic progression and has at least $Ne^{-c\sqrt{\log N}}$ points. ✠

4. Convolution

For functions $f, g \in \mathbb{C}^G$ we define their **convolution** $f * g$ by letting

$$(f * g)(z) := \mathbb{E}_{x \in G}[f(x)g(z-x)] = \mathbb{E}_{x,y}[f(x)g(y) \mid x + y = z].$$

Here we extend our previous notation of $\mathbb{E}$: we define $\mathbb{E}_x[f(x) \mid A]$ to be the average of $f(x)$ over all x that satisfy A. The definition of convolution may be a bit confusing at first, so let us look at a few important examples.

- For an $S \subseteq G$ consider the characteristic function 1_S on G. Then for any $f \in \mathbb{C}^G$ we have

  $$(f * 1_S)(z) = \mathbb{E}_x[f(x)1_S(z-x)] = \frac{1}{|G|} \sum_{x \in z-S} f(x).$$

  So convolution with a characteristic function provides an averaging operator on functions. It is easy to imagine how useful this is in real-world applications, where all data is noisy, so we are not so much interested in the (measured) function f, as in some smoothed-out version of it. This suggests a useful choice for S is a "ball around 0"—a set $\{-t, \ldots, t\}$ in $\mathbb{Z}_n$, say. Even better results can be achieved by replacing 1_S with another function, appropriately concentrated around 0. We will not pursue this direction here. In Section 6 we will meet a variant of it in the form of a "noise operator."

- Consider sets $A, B \subseteq G$. Then

$$\begin{aligned}(1_A * 1_B)(x) &= \mathbb{E}_{a,b}[1_A(a)1_B(b) \mid a + b = x] \\ &= \frac{\text{the number of solutions to } a + b = x}{|G|^2}.\end{aligned}$$

In particular, $\mathrm{supp}(1_A * 1_B) = A + B$. Here $A + B = \{a + b : a \in A, b \in B\}$ and the support was defined in Exercise 2.12. This is frequently used in arithmetic combinatorics to study properties of integer addition. As an example, the proof of Theorem 3.2 we gave in the last section can be modified to use convolutions and get to Equation (11) in a different way. Compare also Exercise 4.3.

- For a polynomial $f(x) = \sum_{i=0}^{n} a_i x^i$ let us use F to denote the mapping $\mathbb{Z}_{2n+1} \to \mathbb{C}$ that maps k to a_k and numbers $n+1, \ldots, 2n$ to 0. Let us now consider two such polynomials, $f(x)$ and $g(x)$, together with their product $h(x) = f(x)g(x)$. It is easy to verify that

$$H = (2n+1) \cdot (F * G). \tag{13}$$

We will return to this soon with a practical application.

Exercise 4.1. Verify that for functions f, g, h defined on a finite abelian group and for complex numbers a, b the convolution satisfies the following:

(a) $f * g = g * f$;

(b) $f * (g * h) = (f * g) * h$; and

(c) $(af + bg) * h = a(f * h) + b(g * h)$. ✠

Exercise 4.2. Let G be a finite abelian group and $f, g \colon G \to \mathbb{C}$. Prove that $\mathrm{supp}(f * g) \subseteq \mathrm{supp}(f) + \mathrm{supp}(g)$. ✠

Exercise 4.3. Let G be an abelian group with n elements. Suppose $p : G \to [0, \infty)$ satisfies $\mathbb{E}_x[p(x)] = 1$. We interpret p as a probability distribution of a random variable X with values in G, in the sense that $\mathrm{Prob}[X = x] = p(x)/n$ for every $x \in G$. Let Y be another random variable with the same distribution, so that X and Y are independent. Show that $X + Y$ has distribution $p * p$. ✠

Convolution as multiplication. The following property is one of the miracles that make the Fourier transform so useful: it translates a useful operation (convolution) to one that is easy to understand and compute (pointwise multiplication).

For $f, g\colon G \to \mathbb{C}$ we have

$$\widehat{f * g} = \widehat{f} \cdot \widehat{g}. \tag{14}$$

The proof is extremely straightforward, though:

$$\begin{aligned}
\widehat{f * g}(a) &= \mathbb{E}_{y\in G}[(f * g)(y)\overline{\chi_a(y)}] \\
&= \mathbb{E}_{y\in G}\left[\mathbb{E}_{x\in G}[f(x)g(y-x)\overline{\chi_a(x)}\ \overline{\chi_a(y-x)}]\right] \\
&= \mathbb{E}_{x\in G}\left[\mathbb{E}_{z\in G}[f(x)g(z)\overline{\chi_a(x)}\ \overline{\chi_a(z)}]\right] \\
&= \widehat{f}(a)\widehat{g}(a).
\end{aligned}$$

Exercise 4.4. Prove the following converse to Equation (14): For $f, g\colon G \to \mathbb{C}$

$$\widehat{fg}(a) = \sum_{b\in G} \widehat{f}(a-b)\widehat{g}(b).$$

We remark that the sum above resembles the definition of convolution $(\widehat{f} * \widehat{g})(a)$. However, as in the case of the inner product and the norm, we are missing the scaling factor $1/|G|$. ✠

Fast multiplication of polynomials. Suppose we wish to multiply two polynomials of degree at most n, $f(x)$ and $g(x)$. There is an obvious algorithm taking $O(n^2)$ steps and it is almost clear that there is no faster way. Well, it turns out there is a faster way, if a bit counter-intuitive.

As above, in the lines preceding Equation (13), we define functions $F, G, H\colon \mathbb{Z}_{2n+1} \to \mathbb{C}$ that encode the sequence of coefficients of $f(x)$, $g(x)$, and their product. We have already argued that $H = (2n+1)\cdot(F * G)$. It still takes the same $O(n^2)$ steps to compute that directly. However, using FFT we can find $\widehat{F}, \widehat{G}$ in time $O(n \log n)$. Equation (14) implies that $\widehat{H}/(2n+1)$ is simply the pointwise product of $\widehat{F}$ and $\widehat{G}$, so we can compute $\widehat{H}$ in time $O(n)$. Another application of FFT recovers H from $\widehat{H}$. Putting it all together, we computed the product $f(x)g(x)$ in time $O(n \log n)$. (We leave aside possible numerical issues though.)

This technique can also be used for multiplication of large integers. Indeed, an integer with base b representation $(a_n \dots a_0)_b$ can

be understood as an evaluation of a polynomial $a(x) = \sum_{k=0}^{n} a_k x^k$ at $x = b$. In fact, when dealing with a number written in binary, we can easily take $b = 2^\ell$ for any $\ell \geq 1$, treating ℓ-tuples of binary digits as one base b digit. We point the reader interested in the details, including what is the optimal choice of ℓ and how much precision is needed, to Knuth's treatise [**Knu73**], Section 4.3.3.C.

We will conclude this section with a useful result on convolution. We skip the proof, but invite the reader to verify two special cases in the exercise below.

Theorem 4.5 (Young's inequality for convolution)**.** *Let f, g be functions and let $p, q, r \in [1, \infty]$ satisfy $\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1$. Then*

$$\|f * g\|_r \leq \|f\|_p \|g\|_q.$$

Exercise 4.6. Verify the following inequalities for functions f, g defined on an abelian group.

(a) $\|f * g\|_1 \leq \|f\|_1 \cdot \|g\|_1$.

(b) $\|f * g\|_\infty \leq \|f\|_p \cdot \|g\|_q$, where $1/p + 1/q = 1$. ✠

5. Poisson Summation Formula

We start with a motivating example. Let $G = \mathbb{Z}_k \oplus \mathbb{Z}_\ell$ and consider characteristic function 1_S of a set $S = \{(x, 0) : x \in \mathbb{Z}_k\}$. We may also write this function explicitly using the Kronecker delta function

$$1_S(x, y) = \delta_0(y).$$

Let us compute the Fourier coefficients. By definition,

$$\begin{aligned}\widehat{1_S}(a, b) &= \mathbb{E}_{(x,y)}[1_S(x, y)\overline{\chi_{a,b}(x, y)}] \\ &= \mathbb{E}_{(x,y)}[\delta_0(y)\overline{\chi_a(x)\chi_b(y)}] \\ &= \tfrac{1}{\ell}\, \mathbb{E}_{x \in \mathbb{Z}_k}[\overline{\chi_a(x)}] \\ &= \tfrac{1}{\ell}\delta_0(a).\end{aligned}$$

Thus, the Fourier transform of 1_S is a multiple of the characteristic function of another set, namely $\{(0, y) : y \in \mathbb{Z}_\ell\}$. Similar result, for a group $\mathbb{Z}_n$ with $n = k\ell$, appeared earlier in Exercise 2.3. Now we present a general framework that includes this phenomenon as Exercise 5.2.

Let S be a subset of a (finite abelian) group G. Its *orthogonal complement* is the set[11]

$$S^\perp = \{a \in G : \chi_a(x) = 1 \text{ for every } x \in S\}.$$

It is easy to observe that $S^\perp$ is a subgroup of G. Another easy (though important) fact is that

$$(H^\perp)^\perp = H$$

whenever H is a subgroup of G. We will need the following version of Lemma 1.4.

Lemma 5.1. *Let G be a finite abelian group, let H be a subgroup of G. Then*

$$\mathbb{E}_{x\in H}[\chi_a(x)] = \begin{cases} 1 & \text{if } a \in H^\perp \\ 0 & \text{otherwise.} \end{cases}$$

Proof. The result is obvious for $a \in H^\perp$, so assume the contrary. Note that χ_a, or rather its restriction to H, is a character of H. As $a \notin H^\perp$, this restriction is a nontrivial character of H, thus an application of Lemma 1.4 finishes the proof. □

Exercise 5.2. Let G be a finite abelian group and let H be a subgroup of G. Then

$$\widehat{1_H}(a) = \begin{cases} \frac{|H|}{|G|} & \text{if } a \in H^\perp; \\ 0 & \text{otherwise.} \end{cases}$$ ✠

The following result, although easy to prove in our setting, is an important fact with far-reaching generalizations.

[11] The name and the notation suggest that this is related to the notion of orthogonal complement in vector spaces. Indeed, when G is the vector space $\mathbb{Z}_p^k$ for a prime p, then $S^\perp$ is the vector-space orthogonal complement defined using the bilinear form $(a,x) \mapsto \sum_{i=1}^k a_i x_i$. For a general group, the definition of $S^\perp$ is related to that of *polar lattice*. We will not pursue this connection, though, for the sake of concreteness.

Theorem 5.3 (Poisson summation formula). *Let G be a finite abelian group and H a subgroup of G. Consider $f\colon G \to \mathbb{C}$ and $x \in G$. Then*

$$\frac{1}{|H|}\sum_{y\in H} f(x+y) = \sum_{a\in H^\perp} \widehat{f}(a)\chi_a(x).$$

Exercise 5.4. Prove Theorem 5.3 (using Lemma 5.1 can be helpful). ✠

Exercise 5.5. Let G be a finite abelian group and H a subgroup of G. Apply the Poisson summation formula to prove that

$$|G| = |H| \cdot |H^\perp|.$$

In particular, we get that $|H|$ divides $|G|$, which is Lagrange's theorem from basic group theory. (The theorem also holds for nonabelian groups, though.) ✠

It is instructive to put extreme values of H in Theorem 5.3.

- If $H = \{0\}$ is the trivial group, then $H^\perp = G$ and we obtain the basic inversion formula

$$f(a) = \sum_{y\in G} \widehat{f}(y)\chi_y(a).$$

- If $H = G$ then $H^\perp = \{0\}$ and we get the often-used result

$$\mathbb{E}_x[f(x)] = \widehat{f}(0).$$

Gauss sum. One of the basic properties of characters is that the average over the whole group is zero (we exclude the trivial character here). For the cyclic group $\mathbb{Z}_p$ this amounts just to a sum of a geometric series $1 + \omega + \cdots + \omega^{p-1}$ for ω a p-th root of unity. From this point of view it may be surprising that one may say meaningful things about such sum over just a subset of $\mathbb{Z}_p$, namely over the set of *quadratic residues.*[12]

[12] We say an $x \in \mathbb{Z}_p$ is a quadratic residue if $x = y^2$ for some $y \in \mathbb{Z}_p$. Some people also use the shorter term (perfect) square instead of the traditional notion quadratic residue.

Let us define the *Gauss sum*

$$\operatorname{Gau}(r) = \sum_{x \in \mathbb{Z}_p} e(rx^2/p).$$

It took Gauss several years to prove the remarkable fact that $\operatorname{Gau}(1) = \sqrt{p}$ for primes $p \equiv 1 \pmod 4$ and $\operatorname{Gau}(1) = i\sqrt{p}$ for primes $p \equiv 3 \pmod 4$. In the next exercise we prove part of this result, up to determining the sign—which is the hardest part, actually.

Exercise 5.6. Let p be a prime.

(a) Prove that $\operatorname{Gau}(rs^2) = \operatorname{Gau}(r)$ for $s \in \mathbb{Z}_p \setminus \{0\}$.

(b) Verify that if -1 is not a quadratic residue in $\mathbb{Z}_p$ then $\operatorname{Gau}(-r) = -\operatorname{Gau}(r)$.

(c) Let $G = \mathbb{Z}_p^2$ and define $f \colon G \to \mathbb{C}$ by putting $f(x_1, x_2) = e(r(x_1^2 + x_2^2)/p)$. Finally, let $H = \mathbb{Z}_p \oplus \{0\}$ be a subgroup of G. Apply Theorem 5.3 to prove that $\operatorname{Gau}(1)^2 = \pm p$. (The sign depends on whether -1 is a quadratic residue.) ✠

Exercise 5.7. Let p be a prime. For $x \in \mathbb{Z}_p$ we define the *Legendre symbol*

$$\left(\frac{x}{p}\right) = \begin{cases} 0 & \text{if } x = 0, \\ 1 & \text{if } x \neq 0 \text{ and } x \text{ is a quadratic residue}, \\ -1 & \text{otherwise.} \end{cases}$$

We put $h(x) = \left(\frac{x}{p}\right)$. Show that h is an eigenvector of the Fourier tranform; that is

$$\widehat{h}(x) = c \cdot h(x)$$

for every $x \in \mathbb{Z}_p$ with a suitable constant c. Compare with Exercise 7.12. ✠

Before closing this section we point out that while our focus has been mainly on the finite groups, the theory is richer for the infinite groups. The infinite case of the Poisson summation formula appears as Theorem 7.13 below.

6. Influence of Variables

6.1. The KKL Theorem. In this section we will study Boolean functions of n variables. Consider a function $f\colon \{0,1\}^n \to \{0,1\}$. We want to understand how much does each of the variables control the value of f. A natural way to formalize this is to set $\mathrm{Inf}_k(f)$ (the *influence of* x_k *on* f) to be the probability that switching x_k changes the value of f, when the other coordinates are fixed on randomly chosen values. Explicitly,

$$\mathrm{Inf}_k(f) = \mathrm{Prob}_{x\in\mathbb{Z}_2^n}[f(x) \neq f(x+e_k)]$$

where e_k is the vector with a single 1 at the k-th position. We have replaced $\{0,1\}^n$ with $\mathbb{Z}_2^n$ to use the group structure. The range of f, on the other hand, will be conveniently understood as a subset of $\mathbb{C}$.

It is obvious that if $f(x) = \sum_{k=1}^n x_k$ (sum in $\mathbb{Z}_2$), then the influence of each variable is 1. For the *dictatorship* $g(x) = x_1$, the first variable has influence 1, every other variable influence 0. In Exercise 6.2 below you are welcome to check that the influence of each variable on the *majority function* is of order $1/\sqrt{n}$.

Hopefully, by now the reader is motivated to find out what can be said about the influence of variables in a general Boolean function. The most natural question seems to be, how small the influence of the "most important" variable can be. In particular, is the majority function the extreme case, as it seems at the first glance? The "sociological implication" of this is that a function with all influences small constitutes a voting system that is robust against bribing (see Exercise 6.6). Kahn, Kalai, and Linial [**KKL88**] proved the following influential result, frequently referred to just by the initials of the authors. Besides its immediate consequences (some of which are provided in the exercise), their result introduced a new proof technique into theoretical computer science.

Theorem 6.1 (KKL theorem)**.** *Let* $f\colon \mathbb{Z}_2^n \to \{0,1\}$ *be a function and put* $\mu = \mathrm{Prob}[f(x)=1]$. *Then for some* k *we have*

$$\mathrm{Inf}_k(f) \geq c\mu(1-\mu)\frac{\log n}{n},$$

where $c > 0$ *is a suitable constant.*

Exercise 6.2. (a) Determine the influence of the majority function: for an odd n the function $\mathrm{Maj}(x_1,\ldots,x_n)\colon \{0,1\}^n \to \{0,1\}$ is defined as the more frequent value among $x_1,\ldots,x_n$.

(b) Using a formula in disjunctive normal form, construct an example of $f\colon \mathbb{Z}_2^n \to \{0,1\}$ with $\mathrm{Inf}_k(f) = \frac{2\ln n}{n}(1+o(1))$ for every k. This implies Theorem 6.1 is tight, up to the choice of c. (Note that we use $\ln n$, the natural logarithm, to get the right constant.) ✠

Proof. Let $\mathrm{Inf}(f)$ denote the *total influence* of a function: $\mathrm{Inf}(f) = \sum_{k=1}^n \mathrm{Inf}_k(f)$. We will actually prove a more general statement, which reads as follows:

(15) For every $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$, we have $\mathrm{Inf}(f) \geq c\mu(1-\mu)\log\frac{1}{\delta}$, where $\delta := \max_k \mathrm{Inf}_k(f)$.

(We leave the derivation of the KKL theorem itself from this to the reader as Exercise 6.3.)

We start by defining $\partial_k f(x) = f(x+e_k) - f(x)$ (a formal analogue of the partial derivative) and noting that

$$\mathrm{Inf}_k(f) = \mathbb{E}_x\left[|\partial_k f(x)|^p\right] = \|\partial_k f\|_p^p. \tag{16}$$

(We are using the fact that the values of $\partial_k f$ are in $\{0,-1,1\}$.) This result holds for every $p \geq 1$. We will need it for two different values of p; one of them is, not surprisingly, $p = 2$.

The first impulse (definitely at this point of this chapter) should be to find the Fourier coefficients of $\partial_k f$ and use Parseval's theorem. An application of Exercise 2.10 gives us $\widehat{\partial_k f}(a) = \chi_a(e_k)\widehat{f}(a) - \widehat{f}(a)$. This equals 0 if $a_k = 0$ and to $-2\widehat{f}(a)$ otherwise. Putting all together, we obtain

$$\begin{aligned}\sum_k \mathrm{Inf}_k(f) &= \sum_k \|\widehat{\partial_k f}\|_2^2 = \sum_k \sum_{a:a_k=1} |\widehat{\partial_k f}(a)|^2\\ &= \sum_k \sum_{a:a_k=1} (-2\widehat{f}(a))^2 = \sum_a \sum_{k:a_k=1} (2\widehat{f}(a))^2.\end{aligned}$$

We thus arrive at

$$\mathrm{Inf}(f) = \sum_{k=1}^n \mathrm{Inf}_k(f) = 4\sum_a |a|\widehat{f}(a)^2, \tag{17}$$

where $|a|$ denotes the number of 1s in a. Here and later on, we use the fact that $\widehat{f}(a)$ is a real number, otherwise we would have to write $|\widehat{f}(a)|^2$.

Up until now, our usage of the Fourier transform was bijective in nature: we have simply used it to express a function in a different basis. Equation (17) breaks new ground. We multiply $\widehat{f}(a)^2$ by $|a|$, thus treating high frequencies (those with large $|a|$) differently. Unsurprisingly, this calls for a new tool. To be more precise, we define (for a fixed integer t)

$$f_{\text{low}} := \sum_{a:|a|\leq t} \widehat{f}(a)\chi_a \tag{18}$$

as the "low-degree part" of f; that is, f_{low} is obtained from f by truncating the Fourier expansion at level t. We observe that this operation plays nicely with the operator ∂_k defined above; we have

$$\partial_k(f_{\text{low}}) = (\partial_k f)_{\text{low}},$$

which is encouraging.

The rough plan is as follows: When f contains a lot of weight on high frequencies, the influences are large. (Either because the χ_a change a lot when $|a|$ is large, or, equivalently, using Equation (17).) Otherwise, f_{low} is large. The missing tool is Proposition 6.8 below that bounds $\|f_{\text{low}}\|_2$ using $\|f\|_{4/3}$. In the next section we explain why this type of bound is reasonable to expect (and we prove it, too). Now, we will combine all of our tools to prove the KKL theorem.

We start slowly, with a quick Fourier-analytic proof of the following inequality for the total influence of a Boolean function in terms of the *variance* $\operatorname{Var} f = \mathbb{E}[f^2] - \mathbb{E}[f]^2$:

$$\operatorname{Inf}(f) \geq 4 \operatorname{Var} f = 4\mu(1-\mu). \tag{19}$$

Using Equation (17) derived earlier and Parseval's identity, we obtain

$$\tfrac{1}{4}\operatorname{Inf}(f) \geq \sum_{a:a\neq 0} \widehat{f}(a)^2 = \|f\|_2^2 - \widehat{f}(0)^2 = \mathbb{E}[f^2] - \mathbb{E}[f]^2 = \operatorname{Var} f.$$

To get (15) we need to gain an extra logarithmic factor. We will use Equation (17) in a similar spirit, but with an extra twist. Let us

write $W := \sum_{a:a\neq 0} \widehat{f}(a)^2 = \mu(1-\mu)$ (the last equality is from the short proof above). We distinguish two cases, depending on whether f has more weight at "high" or "low" Fourier coefficients. We fix the threshold

$$t := \lfloor \tfrac{1}{2} c \log \tfrac{1}{\delta} \rfloor$$

to separate low from high.

Case 1: main weight at high coefficients. Here we assume

$$\sum_{a:|a|>t} \widehat{f}(a)^2 \geq \tfrac{W}{2}.$$

Then we are done quickly:

$$\begin{aligned} \mathrm{Inf}(f) &= 4\sum_a |a|\widehat{f}(a)^2 \geq 4(t+1) \sum_{a:|a|>t} \widehat{f}(a)^2 \\ &\geq 2(t+1)W \geq c\mu(1-\mu)\log\tfrac{1}{\delta} \end{aligned}$$

(here we see where the value of t comes from). This is the easy part of our plan, "high frequencies imply large influences."

Case 2: main weight at low coefficients. This is the complement of Case 1; i.e., now $\sum_{a:0<|a|\leq t} \widehat{f}(a)^2 > W/2$.

Here we use the assumption that $\delta = \max_k \mathrm{Inf}_k(f)$ (so far we have not needed it), and we show that $\mathrm{Inf}(f)$ is even larger than claimed in the theorem.

For a while, we will work on an individual influence $\mathrm{Inf}_k(f)$ for k fixed. Let $g := \partial_k f$; we will apply the (squared) hypercontractive inequality from Proposition 6.8 to g. For simplicity, we use the specific numerical values $C = \sqrt{3}$ and $p = \frac{4}{3}$ obtained in the proof of the proposition. Thus

$$\|g_{\mathrm{low}}\|_2^2 \leq 3^t \|g\|_{4/3}^2 = 3^t \big(\|g\|_{4/3}^{4/3}\big)^{3/2} = 3^t\,\mathrm{Inf}_k(f)^{3/2},$$

(we have used Equation (16) for the second time, now with $p = 4/3$).

Roughly speaking, this tells us that if the influence $\mathrm{Inf}_k(f)$ is small, say smaller than 3^{-2t}, then the contribution of the low Fourier coefficients to it is even considerably smaller.

Now $3^t = 3^{c\log(1/\delta)} \leq \delta^{-1/4}$, say, for c sufficiently small. We estimate $\mathrm{Inf}_k(f)^{3/2} \leq \mathrm{Inf}_k(f)\cdot\delta^{1/2}$, then we sum over k, and the

total influence will appear on the right-hand side of the resulting inequality:

$$\sum_{k=1}^{n} \|(\partial_k f)_{\text{low}}\|_2^2 \le 3^t \delta^{1/2} \sum_{k=1}^{n} \operatorname{Inf}_k(f) \le \delta^{1/4} \operatorname{Inf}(f).$$

In the same way as when we expressed $\operatorname{Inf}(f)$ in terms of the Fourier coefficients (the lines preceding (17)), we get that the left-hand side equals

$$4 \sum_{a:|a|\le t} |a| \widehat{f}(a)^2 \ge 4 \sum_{a:0<|a|\le t} \widehat{f}(a)^2 \ge 2W.$$

Thus

$$\operatorname{Inf}(f) \ge 2W\delta^{-1/4} = 2\mu(1-\mu)\delta^{-1/4} \ge c\mu(1-\mu)\log\tfrac{1}{\delta},$$

where the last inequality holds easily for small enough c. Theorem 6.1 is proved—we still need to prove the new tool, Proposition 6.8, though. □

Exercise 6.3. Explain how Theorem 6.1 follows from the auxiliary statement (15) (possibly with a different value of c). ✠

Exercise 6.4. We say a Boolean function f is *monotone* if we have $f(x) \le f(y)$ whenever $x_k \le y_k$ is true for each k.

(a) If $f\colon \{0,1\}^n \to \{0,1\}$ is monotone, then $\operatorname{Inf}(f) = O(\sqrt{n})$.

(b) The bound in part (a) is asymptotically tight. ✠

Exercise 6.5. (a) Interpret Equation (19) as a type of isoperimetric inequality: a bound for the smallest number of edges connecting a subset S of vertices of the Hamming cube Q_n to $\bar{S}$, the complement of S. (We have defined Q_n in Exercise 1.8 above.)

(b) Prove the following result (usually called the edge isoperimetric inequality for the cube): For every set $S \subseteq V(Q_n)$, the number of edges connecting S and $\bar{S}$ is at least $|S| \log_2 \frac{2^n}{|S|}$.

[Hint: use induction on n.]

(c) Find cases where the inequality in part (b) is tight.

(d) Express the inequality in part (b) as a result about the total influence of a Boolean function. Compare its strength with Theorem 6.1. ✠

A different treatment of the isoperimetric inequality (for the Hamming cube and for other instances) is given in Chapter 2.

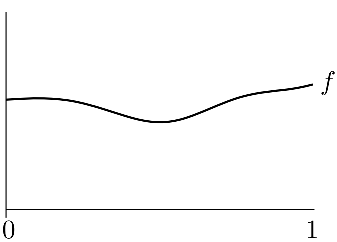
Exercise 6.6. Let $f\colon \{0,1\}^n \to \{0,1\}$ be a monotone Boolean function. Prove that there is for each $\varepsilon > 0$ a set $S \subset [n]$ of size at most $c_\varepsilon n/\log n$ such that

$$\text{Prob}[f(x) = 1 \mid x_k = 1 \text{ for each } k \in S] = 1 - \varepsilon.$$

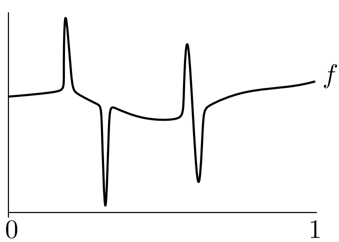
(The constant c_ε depends on ε but not on n.) Thus, if we understand f as a voting scheme, then the voters in S almost surely determine the decision (value of f), provided the other voters choose randomly. ✠

6.2. Lp Norms and a Hypercontractive Inequality. Recall that in Section 1 we defined the L^p norm $\|f\|_p = \big(\mathbb{E}[|f|^p]\big)^{1/p}$ for every $p \geq 1$. Until the last section, however, we have been using only the L^2 norm in applications. Of course, this is because the Parseval equality uses exactly this norm. Now that we have seen (in the proof of Theorem 6.1) a convincing application of another norm, it is perhaps a good time to discuss the hierarchy of all L^p norms.

One can say that for p small (close to 1), the L^p norm measures mainly the "typical" behavior of f (in particular, $\|f\|_1 = \mathbb{E}[|f|]$), while for larger p, the value of $\|f\|_p$ is more influenced by "spikes" of f. Here is a pictorial analog for functions $[0,1] \to \mathbb{R}$.



$\|f\|_p$ and $\|f\|_q$ not very different



$\|f\|_q$ much larger than $\|f\|_p$ for $p < q$

Exercise 6.7. (a) For every function f and $1 \leq p \leq q$ we have

$$\|f\|_p \leq \|f\|_q.$$

(b) Let $\hat{\|}x\hat{\|}_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ be the usual ℓ^p norm defined on $\mathbb{R}^n$. Suppose again that $1 \le p \le q$. Then we have the reverse inequality

$$\hat{\|}x\hat{\|}_p \ge \hat{\|}x\hat{\|}_q.$$ ✠

Let us define the **degree** of f as

$$\max\{|a| : \widehat{f}(a) \neq 0\},$$

the maximum "level" of a nonzero Fourier coefficient.[13]

For a function $f\colon \mathbb{Z}_2 \to \mathbb{C}$ and a threshold parameter $t \ge 0$ we define f_{low}, the "low-degree part" of f, by formula (18). We come to the main result of this section.

Proposition 6.8 (A hypercontractive inequality)**.** *There are constants C and $p < 2$ such that for every $f\colon \{0,1\}^n \to \mathbb{R}$ and every t we have*

$$\|f_{\mathrm{low}}\|_2 \le C^t \|f\|_p.$$

We will comment on the meaning of the word "hypercontractive" at the end of this section. Here we just observe that the L^2 norm on the left-hand side is more sensitive to spikes than the L^p norm on the right-hand side. So we can think of the inequality as a quantitative expression of the intuitive fact that removing high-level components of the Fourier expansion makes f smoother.

We will prove the proposition with $p = 4/3$ and $C = \sqrt{3}$. The main step in the proof is the following lemma.

Lemma 6.9. *Let $f\colon \mathbb{Z}_2^n \to \mathbb{R}$ be a function of degree at most t. Then*

$$\|f\|_4 \le (\sqrt{3})^t \|f\|_2.$$

Proof. The exponents 2 and 4 are convenient, since they allow for a relatively simple inductive proof. We actually prove the fourth power

[13] We can formally express f as a (multilinear) polynomial. To this end, we write $\chi_a(u) = \prod_{i:a_i=1}(1 - 2u_i)$, which is a polynomial of degree $|a|$ (note that this involves a "type cast": while we usually add the u_i modulo 2, here we regard them as real numbers). Then the degree of f is the degree of the corresponding polynomial.

If we wrote the Hamming cube as $\{-1,1\}^n$ (with coordinate-wise multiplication as the operation), then $\chi_a(u)$ would simply be the monomial $\prod_{i:a_i=-1} u_i$.

of the required inequality, i.e.,

$$\|f\|_4^4 \leq 9^t \|f\|_2^4$$

by induction on n.

In the inductive step, we want to get rid of the last variable u_n. We split the Fourier expansion of f into two parts, one with the characters that do not depend on u_n, and the other with those that do. For $u \in \{0,1\}^n$, let $\overline{u} := (u_1, \ldots, u_{n-1})$, and for $b \in \{0,1\}^{n-1}$, we write $b0$ for $(b_1, b_2, \ldots, b_{n-1}, 0)$ (and similarly for $b1$). We have

$$\begin{aligned} f(u) &= \sum_{a \in \{0,1\}^n} \widehat{f}(a)\chi_a(u) \\ &= \sum_{b \in \{0,1\}^{n-1}} \widehat{f}(b0)\chi_{b0}(u) + \sum_{b \in \{0,1\}^{n-1}} \widehat{f}(b1)\chi_{b1}(u) \\ &= \sum_{b \in \{0,1\}^{n-1}} \widehat{f}(b0)\chi_b(\overline{u}) + (-1)^{u_n} \sum_{b \in \{0,1\}^{n-1}} \widehat{f}(b1)\chi_b(\overline{u}) \\ &= g(\overline{u}) + (-1)^{u_n} h(\overline{u}). \end{aligned}$$

We put $g(\overline{u})$ and $h(\overline{u})$ to be the first and the second sum in the second to last line. Observe that g is of degree at most t and h of degree at most $t-1$; we will need it soon. By the orthogonality of the characters we can see that

$$\|f\|_2^2 = \|g\|_2^2 + \|h\|_2^2.$$

We can begin the calculation for the inductive step.

$$\begin{aligned} \|f\|_4^4 &= 2^{-n} \sum_{u \in \{0,1\}^n} f(u)^4 \\ &= 2^{-n} \left[\sum_{v \in \{0,1\}^{n-1}} (g(v) + h(v))^4 + \sum_{v \in \{0,1\}^{n-1}} (g(v) - h(v))^4 \right]. \end{aligned}$$

We expand the fourth powers according to the binomial theorem; the terms with odd powers cancel out, while those with even powers appear twice, and we arrive at

$$\begin{aligned} \|f\|_4^4 &= 2 \cdot 2^{-n} \sum_v \left(g(v)^4 + 6g(v)^2 h(v)^2 + h(v)^4 \right) \\ &= \|g\|_4^4 + 6\langle g^2, h^2 \rangle + \|h\|_4^4 \end{aligned}$$

(the norms and the inner product are now for functions on $\{0,1\}^{n-1}$, one dimension less).

For the terms $\|g\|_4^4$ and $\|h\|_4^4$ we will simply use the inductive assumption. The only trick is with estimating the inner product $\langle g^2, h^2\rangle$: for that we use the Cauchy–Schwarz inequality $\langle x, y\rangle \le \|x\|_2\|y\|_2$, which in our case gives $\langle g^2, h^2\rangle \le \|g^2\|_2\|h^2\|_2 = \|g\|_4^2\|h\|_4^2$. Only after that we apply induction, and we obtain

$$\begin{aligned}
\|f\|_4^4 &\le \|g\|_4^4 + 6\|g\|_4^2\|h\|_4^2 + \|h\|_4^4 \\
&\le 9^t\|g\|_2^4 + 6\cdot 9^{t/2}\|g\|_2^2 \cdot 9^{(t-1)/2}\|h\|_2^2 + 9^{t-1}\|h\|_2^4 \\
&= 9^t\left(\|g\|_2^4 + 6\cdot 9^{-1/2}\|g\|_4^2\|h\|_4^2 + 9^{-1}\|h\|_2^4\right) \\
&\le 9^t\left(\|g\|_2^2 + \|h\|_2^2\right)^2 \\
&= 9^t\|f\|_2^4.
\end{aligned}$$

The lemma is proved. (Well, wait a second ... what is the basis of the induction?) □

Proof of Proposition 6.8. If we apply the lemma that we have just proved to $f_{\rm low}$, we get

$$\|f_{\rm low}\|_4 \le (\sqrt{3})^t\|f_{\rm low}\|_2, \tag{20}$$

but how do we relate this to the norm of f itself?

The first trick of the proof is to consider the inner product $\langle f_{\rm low}, f\rangle$ and express it using the Fourier coefficients:

$$\langle f_{\rm low}, f\rangle = \sum_a \widehat{f_{\rm low}}(a)\widehat{f}(a) = \sum_{a:|a|\le t}\widehat{f}(a)^2 = \|f_{\rm low}\|_2^2$$

(using Parseval).

The second and last trick is to use a standard tool called Hölder's inequality. For $\langle f_{\rm low}, f\rangle$, with exponents $p = 4$ and $q = \frac{4}{3}$, it gives

$$\|f_{\rm low}\|_2^2 = \langle f_{\rm low}, f\rangle \le \|f_{\rm low}\|_4\|f\|_{4/3}.$$

Now we bound $\|f_{\rm low}\|_4$ using (20), divide the resulting inequality by $\|f_{\rm low}\|_2$, and we arrive at the desired inequality $\|f_{\rm low}\|_2 \le (\sqrt{3})^t\|f\|_{4/3}$. □

On hypercontractive inequalities. Let $(Z, \|.\|)$ be a normed space (that is, Z is a vector space and $\|.\|$ is a norm on Z). A 1-Lipschitz mapping $Z \to Z$ is often called *contractive.* This term is most often used for *linear* mappings $A\colon Z \to Z$, which in this context are referred to as *(linear) operators.* For a linear operator, contractivity means $\|Ax\| \leq \|x\|$ for every $x \in Z$.

Now let us consider *two* different norms $\|.\|$ and $\|.\|'$ on Z, and assume that $\|x\| \leq \|x\|'$ for all x. If the linear operator A even satisfies $\|Ax\|' \leq \|x\|$, it is called **hypercontractive** (it is "more than contractive", since if x is small under the smaller norm, Ax is small even under the bigger norm).

For us, the relevant setting is to let Z be the space of all functions $\{0,1\}^n \to \mathbb{R}$, and $\|.\| = \|.\|_p$, $\|.\|' = \|.\|_q$ with $p < q$ (see Exercise 6.7).

How does Proposition 6.8 fit into this? We regard the "truncation" $f \mapsto f_{\rm low}$ as an operator $L\colon Z \to Z$; it is linear because each Fourier coefficient $\widehat{f}(a)$ depends linearly on f. The proposition then tells us that L is hypercontractive for some $p < 2$ and $q = 2$, well, almost, because there is the factor C^t.

This hypercontractive inequality is relatively easy to prove, and in a sense, it is tailored for the proof of the KKL theorem. The original, and usual, proof of the KKL theorem uses another hypercontractive inequality, proved by Bonami and independently by Gross (often also attributed to Beckner, who proved some generalizations later).

To state it, we introduce, for a real parameter $\rho \in [0,1]$, the **noise operator** $T_\rho\colon Z \to Z$. The simplest way of defining it is in terms of the Fourier expansion:

$$T_\rho f := \sum_a \widehat{f}(a)\rho^{|a|}\chi_a;$$

that is, higher Fourier coefficients are reduced more by T_ρ. In particular, $T_1 f = f$ and $T_0 f$ is the constant function equal to $\widehat{f}(0) = \mathbb{E}[f]$.

To explain the name "noise operator," we need another definition. Let $p := (1-\rho)/2$, and let $x \in \{0,1\}^n$ be a random vector, the noise, where each x_i is set to 1 with probability p and 0 with probability $1-p$, independent of all other x_j's. Then we have

$$T_\rho f(u) := \mathbb{E}_x[f(u+x)].$$

In words, to evaluate the function $T_\rho f$ at some given u, we first flip each coordinate of u with probability p, then we apply f to the resulting vector, and we take the expectation over the random flips. Thus, $T_\rho f(u)$ is a weighted average of the values of f, where (for $\rho < 1$) values at points closer to u are taken with larger weight.

Exercise 6.10. (a) Verify that the two definitions of $T_\rho f$ are equivalent.

(b) Express $T_\rho f$ as a convolution of f with an appropriate function. ✠

Assuming $1 \le p \le q$, the hypercontractive inequality for T_ρ asserts that we have

$$\|T_\rho f\|_q \le \|f\|_p \tag{21}$$

for all $f \in Z$ if (and only if) $\rho^2 \le \frac{p-1}{q-1}$. This, with $q = 2$, was used in the first proof of the KKL theorem. Essentially, one first derives Proposition 6.8 from (21) and then proceeds as we will do in the next section.

In the usual proof of (21), one first proves the case $n = 1$, which is a laborious but essentially straightforward calculus problem. Then one derives the general case from the 1-dimensional one by a general inductive argument; this is often expressed by saying that the inequality (21) *tensors*. This is one of the most common approaches to proving multidimensional inequalities, such as various isoperimetric inequalities: one needs to find a version of the considered inequality that tensors, and then work out the one-dimensional version. (Also see Chapter 2.)

The KKL theorem, as well as (21), can also be derived from a **log-Sobolev inequality** for the Hamming cube, which asserts that

$$\sum_{\{u,v\}\in E} |f(u) - f(v)|^2 \ge \frac{1}{n} \operatorname{Ent}[f^2],$$

where E is the edge set of the Hamming cube, and where the **entropy** of a function g is defined as $\operatorname{Ent}[g] := \mathbb{E}[g \log(g/\mathbb{E}\, g)]$. The log-Sobolev inequality is again proved for the 1-dimensional case and

then tensored. This way of proving the KKL theorem is nicely presented in the survey by Biswal [**Bis11**]. It is probably shorter than the one presented in the current chapter.

7. Infinite Groups

In this section we will extend the theory to infinite groups. We will have to do with a brief introduction to the topic, since the theory is vast, for reasons soon to become clear. The basic idea is still the same: we wish to express a function $f\colon G \to \mathbb{C}$ (defined on an abelian group G) as a certain combination of symmetric functions—characters.

7.1. Characters of Infinite Groups.

A function $\chi\colon G \to \mathbb{T}$ is a *character* of G if it is a group homomorphism (it satisfies the conditions of Definition 1.1) and it is continuous. As before, we let $\widehat{G}$ be the set of all characters of G with pointwise multiplication as the group operation.

Every mapping from a finite set is continuous, so we did not have to add the continuity assumption in Definition 1.1. We will discuss only the three most important instances:

$$\widehat{\mathbb{Z}} \cong \mathbb{T}, \qquad \widehat{\mathbb{R}} \cong \mathbb{R}, \quad \text{and} \quad \widehat{\mathbb{T}} \cong \mathbb{Z}$$

We start gently, by looking for characters of $\mathbb{Z}$. Since $\mathbb{Z}$ is a discrete topological space (see Chapter 6), the continuity condition is trivially satisfied. It is easy to observe that a group homomorphism $f\colon \mathbb{Z} \to \mathbb{T}$ is uniquely defined by $f(1)$. Indeed, if $f(1) = z$ then $f(2) = f(1+1) = z^2$, and in general $f(n) = z^n$ by an easy induction. Thus, we may define χ_z to be the function $n \mapsto z^n$, obviously a character for every $z \in \mathbb{T}$. It is easy to verify that the mapping $z \mapsto \chi_z$ is an isomorphism of $\mathbb{T}$ and $\widehat{\mathbb{Z}}$.

For the other cases we will rely on the following nontrivial fact:

Fact 7.1. *For every continuous group homomorphism* $f\colon \mathbb{R} \to \mathbb{R}/\mathbb{Z}$ *there is a continuous group homomorphism* $g\colon \mathbb{R} \to \mathbb{R}$ *such that* $f(x) = g(x) \bmod 1$.

It remains to understand what mappings $f\colon \mathbb{R} \to \mathbb{R}$ satisfy

$$f(x+y) = f(x) + f(y) \tag{22}$$

for all x, y. (This is easily seen to be equivalent to f being a group homomorphism.) This is the *Cauchy functional equation*, and Cauchy found (as you are encouraged to do in the next exercise) that all continuous solutions are of the form $f(x) = ax$, $a \in \mathbb{R}$. It is worth noting that discontinuous solutions exist, but they are highly nonconstructive; this is the reason why we have continuity in the definition of a character.

Exercise 7.2. Let $f\colon \mathbb{R} \to \mathbb{R}$ be any function satisfying (22).

(a) Prove that $f(q) = qf(1)$ for every rational q.

(b) Conclude that if f is continuous, then $f(x) = ax$ for some $a \in \mathbb{R}$.

(c) Let $\mathcal{R}$ denote the vector space of the reals over the field $\mathbb{Q}$. Show that every function satisfying (22) is a linear mapping from $\mathcal{R}$ to itself. On the other hand, every such linear mapping satisfies (22).

(d) Explain how a linear mapping in part (c) can be constructed[14] using a Hamel basis—a basis of the vector space $\mathcal{R}$. ✠

Using Fact 7.1 and Exercise 7.2, we conclude that every continuous homomorphism $f\colon \mathbb{R} \to \mathbb{R}/\mathbb{Z}$ is of the form $f(x) = ax \bmod 1$, $a \in \mathbb{R}$. Consequently, continuous characters of $\mathbb{R}$ must be of the form $\chi_a(x) = e(ax) = e^{2\pi i a x}$. It is easy to verify that every χ_a is indeed a character of $\mathbb{R}$. We leave the remaining case of $\widehat{\mathbb{T}}$ as a simple exercise.

Exercise 7.3. Find all characters of $\mathbb{T}$. ✠

[14] For the reader who is attentive to the set-theoretical foundations of mathematics, we note that the existence of such basis is guaranteed only when we assume the axiom of choice. In particular, the basis itself cannot be constructed.

Exercise 7.4. Let G be an abelian group and $\widehat{G}$ the group of characters of G. For every $a \in G$ we define mapping $f_a \colon \widehat{G} \to \mathbb{T}$ by $f_a(\chi) = \chi(a)$. Verify that

(a) f_a is a group homomorphism;

(b) the mapping $a \mapsto f_a$ is a group homomorphism. ✠

For space reasons, we have not defined the topology on $\widehat{G}$. Thus, we cannot verify that the function f_a from Exercise 7.4 is continuous. With a proper definition it is true, though, which means that f_a is a character of $\widehat{G}$ for each $a \in G$. The following theorem provides the right extension of Theorem 1.3 to the infinite setting. (The definition of "locally compact" is given in the next section; we omit the definition of the topology on $\widehat{G}$, and thus also on $\widehat{\widehat{G}}$.)

Theorem 7.5 (Pontryagin duality theorem). *Let G be a locally compact abelian group. Then $\widehat{\widehat{G}}$ is isomorphic to G. Explicitly, the mapping $a \mapsto f_a$ is both a group isomorphism and a homeomorphism of topological spaces.*

7.2. Seeking the Right Setting—LCA Groups. We will not attempt much in the general case, but we want to at least describe what the general case is—in what generality can Fourier analysis be meaningfully performed.

An infinite group frequently comes with a topology (see Chapter 6 for an introduction to topology). We will assume the topology is Hausdorff, which is true for all usual groups. We say a topological group G (an object with both topological and group structure) is *locally compact* if the neutral group element 0 has an open neighborhood U such that the closure $\operatorname{cl} U$ is a compact set.

This is trivially true for groups with discrete topology (we can take $U = \{0\}$), as well as for compact groups (where $U = G$ works well). Both of these cases are important (consider groups $\mathbb{T}^n$, $\mathbb{Z}^n$), but the notion of a locally compact group also covers the reals, $\mathbb{R}^n$, and even more exotic groups, such as the p-adic numbers.[15] The main reason why locally compact groups are an important case is that we

[15] We will not discuss p-adic numbers in this book. Let us just mention for the curious reader that they are an extension of rational numbers in a different direction than the reals, with important use in number theory.

can define a measure[16] of its subsets in a reasonable way: so that measures of sets A and $A + x = \{a + x : a \in A\}$ are the same. The reader is advised to read the part about the *Haar measure* in Chapter 1, Section 1.3.

When we have a locally compact group G with Haar[17] measure μ, we can define integrals of functions defined on G.[18] In particular, we can define for a function $f : G \to \mathbb{C}$ the expectation of f in the usual way

$$\mathbb{E}[f] = \int f \, \mathrm{d}\mu. \tag{23}$$

This allows us to define the inner product and the norm as we did in Section 1. When G is a finite group, the Haar measure simply counts sizes: $\mu(A) = |A|/|G|$. Thus, Equation (23) is indeed a proper generalization of the formula for $\mathbb{E}[f]$ given in Section 1 of this chapter.

The definition of the norm and the inner product is useful also if G is not abelian. However, for Fourier analysis on infinite groups commutativity is crucial (compare with Chapter 4 though). It turns out this is enough; from now on we will work with functions defined on a group that is locally compact and abelian (or **LCA** for short).

As in the finite case, we want to define for a function $f : G \to \mathbb{C}$ the *Fourier transform* of f as the function $\widehat{f} : \widehat{G} \to \mathbb{C}$ given by

$$\boxed{\widehat{f}(\chi) := \langle f, \chi \rangle = \mathbb{E}[f\bar{\chi}].}$$

Here comes the first issue: while in Definition 2.2 the sum had a clear sense, the integral hidden in the $\mathbb{E}$ above may be undefined if the function f is not sufficiently nice—measurable. Even for a measurable function the integral may be infinite. To solve both of these issues, we define the Lebesgue space

$$L^1(G) = \Big\{ f : G \to \mathbb{C} : f \text{ is measurable and } \textstyle\int |f| \, \mathrm{d}\mu < \infty \Big\}.$$

---

[16]Think of volume, if you are unfamiliar with measure theory.

[17]The measure is unique up to a multiplicative constant. It is common to choose the constant so that the measure of the whole group is 1, when this is possible—that is, when G is compact.

[18]The reader unfamiliar with the theory of Lebesgue integrals may wish to briefly skim Section 2 of Chapter 1.

Exercise 7.6. Let G be an LCA group, χ a character of G and $f \in L^1(G)$. Show that the integral in the definition of $\widehat{f}(\chi)$ is well-defined (and, in particular, finite). ✠

This is all we will do at the general level. In the next section we will discuss how the inverse Fourier transform translates to the continuous setting for functions defined on $\mathbb{T}$.

7.3. The Case of $\mathbb{T}$: Periodic Functions. In this section we will study the most classical case of the theory: functions defined on $\mathbb{T}$. Using the isomorphism $e\colon \mathbb{R}/\mathbb{Z} \to \mathbb{T}$, this is equivalent to studying 1-periodic[19] functions of a real variable, obviously an important case. Indeed, Fourier series were the starting point of this part of mathematics. We remind the reader that $\mathbb{R}/\mathbb{Z}$ can be identified with $[0,1]$ (or any other interval of length 1); we only need to restrict to functions f satisfying $f(0) = f(1)$.

In the previous section we saw what the characters of $\mathbb{T}$ are. Consequently, the characters of $\mathbb{R}/\mathbb{Z}$ are functions $\chi_n\colon \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ defined by $\chi_n(x) = e^{2\pi i n x} = \cos(2\pi n x) + i \sin(2\pi n x)$. This explains the common terminology: $x \in \mathbb{T}$ is called the *time*, while $n \in \mathbb{Z}$ the *frequency*. People also speak of the *time domain* $\mathbb{T}$ and the *frequency domain* $\widehat{\mathbb{T}} \cong \mathbb{Z}$.

The Haar measure on $\mathbb{R}/\mathbb{Z}$ is just the usual Lebesgue measure,[20] so we have

$$\mathbb{E}[f] = \int_0^1 f(x)\,\mathrm{d}x.$$

For notational convenience we write $\widehat{f}(n)$ for $\widehat{f}(\chi_n)$: thus, we let $\widehat{f}$ be a function defined on $\mathbb{Z}$, instead of the (isomorphic) group $\widehat{\mathbb{T}}$. This gives us the following specialization of the formula for *Fourier transform*. For $f \in L^1$ we define the *Fourier coefficients* of f

[19] Frequently, people prefer to study 2π-periodic functions, motivated by elementary trigonometric functions. The theory is, of course, the same, just some scaling factors need to be added where appropriate.

[20] Strictly speaking, the Haar measure is only defined on the Borel sets, so it is the restriction of the Lebesgue measure to the Borel sets.

$$\widehat{f}(n) = \int_0^1 f(x)e^{-2\pi inx}\,\mathrm{d}x. \tag{24}$$

As before we may want to express f using $\widehat{f}$ (*inverse Fourier transform*). A natural formula to expect is $f(x) = \sum_{n\in\mathbb{Z}} \widehat{f}(n)\chi_n(x) = \sum_{n\in\mathbb{Z}} \widehat{f}(n)e^{2\pi inx}$. However, this sum may be undefined or it may not equal to $f(x)$. Thus, we will write just

$$f(x) \sim \sum_{n\in\mathbb{Z}} \widehat{f}(n)e^{2\pi inx}. \tag{25}$$

The sum in the formula above is the *Fourier series* of f. Next, we will discuss in what sense does the Fourier series determine the value of $f(x)$. Ideally, we would like to have an equality in the above formula, but this is not generally the case. In fact, Kolmogorov constructed an integrable function whose Fourier series diverges at every point!

Sines and cosines vs. exponentials. Before we delve into the fine points of convergence, we explain why we employ exponentials where you may be accustomed to see sines and cosines. Frequently, we deal with real-valued functions and it may be inconvenient to use complex exponentials to express them. Fortunately, it is easy to translate between complex exponentials and combinations of sines and cosines. We will need the following familiar formulas for a real x:

$$e^{ix} = \cos x + i\sin x; \qquad \cos x = \frac{e^{ix} + e^{-ix}}{2}; \qquad \sin x = \frac{e^{ix} - e^{-ix}}{2i}.$$

We will express $\widehat{f}(n)$ as $(a_n - ib_n)/2$ (the scaling factor $\frac{1}{2}$ is used for consistence with the usual formulas). Thus, we get

$$a_n = 2\int_0^1 f(x)\cos 2\pi nx\,\mathrm{d}x, \qquad b_n = 2\int_0^1 f(x)\sin 2\pi nx\,\mathrm{d}x.$$

It is easy to see that $a_{-n} = a_n$ and $b_{-n} = -b_n$ (in particular, $b_0 = 0$). Since we are assuming f is real-valued, a_n and b_n are real numbers, too.

Substituting $\widehat{f}(n) = (a_n - ib_n)/2$ in formula (25) and using the fact that $f(x)$ is real, we get

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty}\Big(\frac{a_n + a_{-n}}{2}\cos 2\pi nx + \frac{b_n - b_{-n}}{2}\sin 2\pi nx\Big).$$

Joining the terms with n and $-n$ yields

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty}\Big(a_n \cos 2\pi nx + b_n \sin 2\pi nx\Big). \tag{26}$$

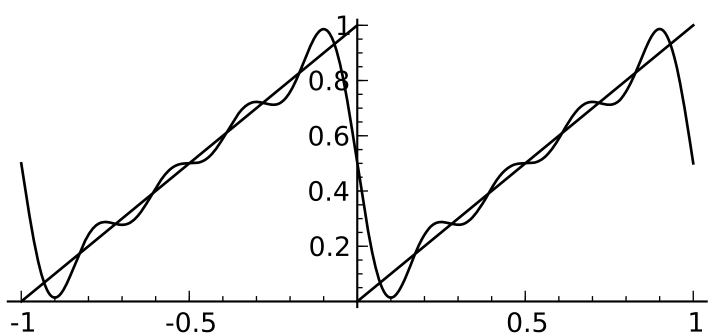
This is the formula that most people recognize as the *Fourier series*, expressing a real-valued 1-periodic function in terms of sines and cosines.

Having convinced ourselves that Equations (25) and (26) are expressing the same statement in different terms, we return to complex exponentials, since the formulas are shorter and the theory easier to present.

Pointwise convergence—Dirichlet's theorem. To start our discussion of the convergence issues of Equation (25) we define

$$S_n(f,x) = \sum_{k=-n}^{n} \widehat{f}(k)e^{2\pi ikx}.$$

As an illustration, here is the function $S_{10}(f,x)$ for $f(x) = x \bmod 1$.



We will assume throughout Section 7.3 that $f \in L^1(\mathbb{R}/\mathbb{Z})$. Recall that a function $f\colon [a,b] \to \mathbb{C}$ is called *piecewise smooth* if the derivative f' exists at all points of $[a,b]$ with finitely many exceptions, where finite left- and right-sided limits of f and of f' exist. In particular, at every x there exists the limit of f from the left $f(x-) = \lim_{y\to x^-} f(y)$, and also $f(x+)$, the limit from the right.

Theorem 7.7 (Dirichlet's theorem)**.** *Let $f\colon \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ be a piecewise smooth function. Then for every x we have*

$$\sum_{k\in\mathbb{Z}} \widehat{f}(k)e^{2\pi ikx} = \frac{f(x-)+f(x+)}{2},$$

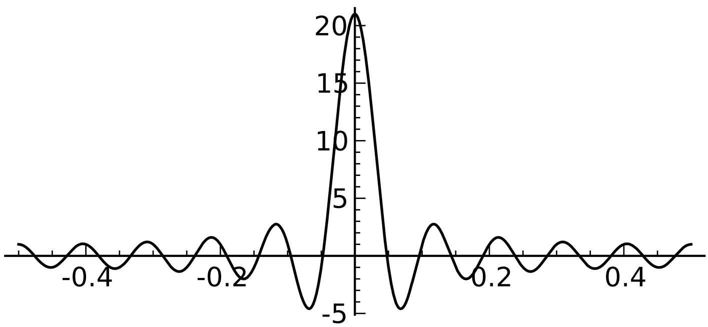
where we have defined the left-hand side to equal the limit of the partial sums, $\lim_{n\to\infty} S_n(f,x)$*. In particular, if f is continuous at x then the sum in formula* (25) *equals $f(x)$.*

If f is continuous at x, then of course the right-hand side in Dirichlet's theorem equals $f(x)$ and the Fourier series sums to $f(x)$. If f is not continuous at x, then we cannot say anything about how the Fourier series relates to $f(x)$—imagine changing a value of a continuous function at a single point, x. However, at a jump discontinuity x it is possible that $f(x)$ is equal $(f(x-)+f(x+))/2$. Thus, we can have a discontinuous function expressed as an infinite sum of complex exponentials (equivalently, sines and cosines). This fact was extremely surprising at the time of Fourier.

To keep the exposition short, we will not prove Theorem 7.7. However, let us mention a key idea of the proof: we can express $S_n(f,x)$ as a convolution $f * D_n$, where the *Dirichlet kernel* $D_n(x)$ is defined as

$$D_n(x) = \sum_{k=-n}^{n} e^{2\pi ikx};$$

here is the graph of D_{10}.



Exercise 7.8. (a) Show that $S_n(f) = f * D_n$.

(b) Prove that

$$D_n(x) = \frac{\sin\big((2n+1)\pi x\big)}{\sin(\pi x)}.$$ ✠

The plot of $D_n(x)$ suggests some of the issues with Dirichlet's theorem: while $D_n(0)$ (or, rather, the limit at 0) equals $2n+1$ (the maximum over $\mathbb{R}/\mathbb{Z}$), there is a lot of oscillation. Indeed, the average value of D_n is only 1, while its L^1-norm, $\mathbb{E}[|D_n|]$, is proportional to $\log n$. For f sufficiently nice, this oscillation turns into a lot of cancellation. This rough idea can be turned into a proof by using the Riemann–Lebesgue lemma (which we will mention shortly at the end of Section 7.3).

Weaker convergence—Fejér's theorem. As a corollary of Dirichlet's theorem we see that smooth functions have nicer convergence properties. While we will see another confirmation of this general fact at the end of this section, frequently we want to understand the behaviour of Fourier series of a non-smooth function. In particular: do the Fourier coefficients $\widehat{f}(k)$ determine the function f? It turns out we can still recover f if we relax our notion of convergence somewhat.

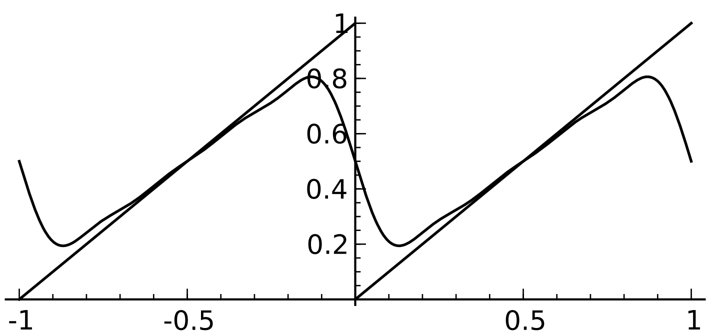
Exercise 7.9. Let $(a_n)_{n=0}^{\infty}$ be a sequence of complex numbers and let (b_n) be the sequence of partial averages (so-called *Cesàro means*):

$$b_n = \frac{1}{n+1} \sum_{k=0}^{n} a_k.$$

(a) Prove that if the sequence a_n converges to a limit a, then the sequence b_n converges to a as well.

(b) Find an example of a sequence that does not converge, but its Cesàro means do. ✠

We apply this general idea to the case of partial sums of the Fourier series. We put $\sigma_n(f,x) = \frac{1}{n+1} \sum_{k=0}^{n} S_k(f,x)$. From Exercise 7.9 above we should expect that as n goes to ∞, the sequence of functions $\sigma_n(f,x)$ behaves better than $S_n(f,x)$. The plot of $\sigma_{10}(f,x)$, again for the function $f(x) = x \bmod 1$, illustrates this.



Theorem 7.10 (Fejér's theorem)**.** *Let $f\colon \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ be a piecewise continuous*[21] *function. Then for every x we have*

$$\lim_{n\to\infty} \sigma_n(f,x) = \frac{f(x-) + f(x+)}{2}$$

whenever the right-hand side exists. If the function is continuous on $\mathbb{R}/\mathbb{Z}$, then the convergence of $\sigma_n(f,x)$ to $f(x)$ is uniform.

We will not prove this result either, but again, the key is to express $\sigma_n(f,x)$ as a convolution. This time we have $\sigma_n(f) = f * F_n$,

[21] That is, f is a function continuous at every point except a finite number of special points, where the limits from the left and from the right exist and are finite.

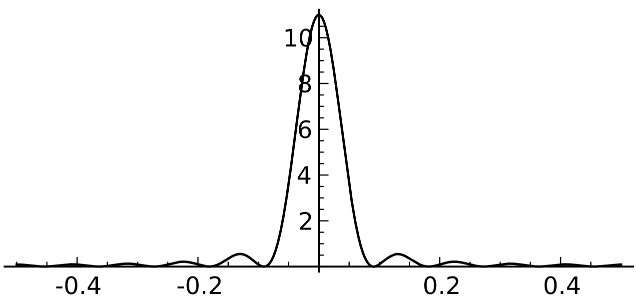
where the *Fejér kernel* $F_n(x)$ is defined by

$$F_n(x) = \frac{1}{n+1} \sum_{k=0}^{n} D_k(x).$$

Exercise 7.11. (a) Show that $\sigma_n(f) = f * F_n$.

(b) Prove that

$$F_n(x) = \frac{1}{n+1} \left(\frac{\sin\big((n+1)\pi x\big)}{\sin(\pi x)} \right)^2 .$$ ✠



The plot of $F_{10}(x)$ above (or the formula for $F_n(x)$, for the analytically-minded readers) explains why the sequence $\sigma_n(f)$ behaves better than $S_n(f)$. First, we have $F_n(x) \geq 0$ for all x, and second, the mass of the function is concentrated around 0. This makes the proof of Fejér's theorem substantially easier than that of Dirichlet's theorem. The nice properties of the Fejér's kernel are generalized by the notion of *approximate identity*.

As an easy corollary of Fejér's theorem we get the following important observation.

> If $f, g\colon \mathbb{T} \to \mathbb{C}$ are continuous functions and $\widehat{f}(r) = \widehat{g}(r)$ for every r, then $f = g$.

An even easier corollary of Fejér's theorem is that any continuous function on $[0, 1]$ can be approximated, with an arbitrary precision, by *some* linear combination of functions $\sin kx$, $\cos kx$, for $k = 0, 1, \ldots, n$. The precision, of course, is determined by how large n

we allow. With some extra work, one can derive the Weierstrass approximation theorem: any continuous function on $[0,1]$ can be approximated with an arbitrary precision by a polynomial.

Of course, both Dirichlet's and Fejér's theorem have stronger statements using, e.g., the assumption that f is of *bounded variation*. This would lead us too far into the intricacies of real analysis, though. As a more modern follow-up to these results, Carleson proved in 1966 that Fourier series of any L^2 function (function with finite $\|f\|_2$) converges almost everywhere.

Many interesting results connect properties of a function with those of its Fourier transform. Classical is the *Riemann–Lebesgue lemma*: for every $f \in L^1(\mathbb{T})$ the Fourier coefficients $|\widehat{f}(n)|$ tend to 0 as $|n|$ goes to ∞. As a refinement of this, if f is k-times continuously differentiable, then $|\widehat{f}(n)| = O((1+|n|)^{-k})$; a version of the converse implication is also true. This is similar to the relation of Dirichlet's and Fejér's theorems: local properties of f (such as behaviour of its derivatives) correspond to global properties of $\widehat{f}$. Further variants could fill volumes, but we have to stop somewhere.

7.4. The Case of $\mathbb{R}^n$. We will be brief in this section. In Section 7.1 we discussed the characters of $\mathbb{R}$. A simple exercise is to conclude that the characters of $\mathbb{R}^n$ are the functions χ_a given for $a \in \mathbb{R}$ by the formula $\chi_a(x) = e^{2\pi i a\cdot x}$; by $a \cdot x$ we mean the usual inner product in $\mathbb{R}^n$, i.e., $a \cdot x = \sum_{k=1}^n a_k x_k$. The Haar measure for $\mathbb{R}^n$ is the usual Lebesgue measure.[22] The formula for the Fourier transform is, as you may have anticipated,

$$\widehat{f}(a) = \langle f, \chi_a \rangle = \int_{\mathbb{R}^n} f(x) e^{-2\pi i a\cdot x}\,\mathrm{d}x.$$

The reader is probably eager to try this formula for a concrete function f. We must be careful though, as $f(x)$ needs to decrease fast with growing $\|x\|$, or the integral will not be defined. An interesting example is given in the next exercise (compare also to Exercise 5.7).

[22] We can also take any multiple of it, there are several competing conventions in the literature. Also, as in Section 7.3, the Haar measure is in fact the restriction of the Lebesgue measure to the Borel sets.

Exercise 7.12. Consider the function of real variable $f(x) = e^{-\pi x^2}$ (i.e., we have $n = 1$). Prove that the Fourier transform satisfies $\widehat{f}(a) = f(a)$ for every $a \in \mathbb{R}$. ✠

The inversion formula is also unsurprising:

$$f(x) \sim \int_{-\infty}^{\infty} \widehat{f}(a) e^{2\pi i a \cdot x} \, \mathrm{d}a.$$

Extensive theory has been developed that answers the obvious question: when can the $\sim$ be replaced by equality? We only mention that it is in particular true (for almost every x) if both f and $\widehat{f}$ are in $L^1(\mathbb{R}^n)$.

Let us leave the topic by mentioning the infinite counterpart to Theorem 5.3.

Theorem 7.13 (Poisson summation formula)**.** *Suppose that $f, \widehat{f} \in L^1(\mathbb{R}^n)$ satisfy*

$$|f(x)| + |\widehat{f}(x)| \leq c(1 + |x|)^{-n-\varepsilon}$$

for some $\varepsilon > 0$ and $c \in \mathbb{R}$. Then for every $x \in \mathbb{R}^n$

$$\sum_{y \in \mathbb{Z}^n} f(x + y) = \sum_{a \in \mathbb{Z}^n} \widehat{f}(a) \chi_a(x).$$

Using this formula, Banaszczyk proved the so-called transference theorem—a result connecting sizes of a lattice in $\mathbb{R}^n$ and its dual. This was used by Lovász, Lenstra, and Schrijver to study the complexity of approximation of the shortest vector in a lattice; see Micciancio and Goldwasser [**MG02**] for details. A more classical application of Theorem 7.13 is the result of Nyquist and Shannon about sampling: suppose that "the highest frequency" of a function $f\colon \mathbb{R} \to \mathbb{R}$ is less than w; that is $\widehat{f}(y) = 0$ if $|y| \geq w$. Then f is uniquely determined by its samples of frequency $2w$; that is, by the values $(f(n/2w))_{n \in \mathbb{Z}}$.

7.5. Fourier Transform and the Derivative. In this short section we explain a little bit of how Fourier analysis is used to solve differential equations—equations that involve an unknown function f and its derivatives f', f'', etc. To this end, we first need to understand the relation between the Fourier transform of a function and that of its derivative. For simplicity, we consider a function for which

Equation (25) holds with equality. If we differentiate it term by term, we obtain that $f'(x) = \sum_{n\in\mathbb{Z}} \widehat{f}(n) 2\pi i n e^{2\pi i n x}$. Equivalently, we get that

$$\widehat{f'}(n) = 2\pi i n \widehat{f}(n)\,. \tag{27}$$

Of course, we must be careful: differentiating an infinite sum term by term is not always justified. Using standard tools from real analysis one can justify this procedure whenever f is three times continuously differentiable. We can do better, though:

Exercise 7.14. Use Equation (24) and integration by parts to verify that Equation (27) holds for every 1-periodic function f with a continuous derivative. ✠

There are many ways to use Fourier analysis to solve differential equations: this is not surprising, as solving the differential equation for heat dissipation was the original motivation of Joseph Fourier. We only illustrate it on a simple example—one that engineers know as driven harmonic oscillator. The idea is simple: if we express both f and g below as Fourier series, then the differential equation becomes a collection of linear equations for the Fourier coefficients.

Exercise 7.15. Use the method described above to find all 1-periodic functions g that satisfy the following differential equation:

$$g''(x) + 2g'(x) + g(x) = f(x),$$

where

(a) $f(x) = \sin 2\pi x$;

(b) $f(x)$ is a 1-periodic function with continuous derivative. ✠

8. Literature

In this chapter we treated the Fourier analysis on finite groups more extensively than the classical theory dealing with infinite groups. The reader who wishes to make up for this and learn a bit more of the classical theory is invited to read Körner's monograph [**Kör89**]; it contains both theory and applications ranging from solving the heat

equation to the proof of the Central Limit Theorem from probability theory (the main idea is alluded to in Exercise 4.3). A quick overview of the classical theory is also provided in the online lecture notes by Tao [**Tao14**].

A book by Terras [**Ter99**] concentrates on the case of finite groups; its second part deals with the nonabelian case, i.e., with the representation theory of finite groups. Many applications ranging from chemistry to error-correcting codes make it an interesting read.

A fruitful area for application of Fourier analysis is additive combinatorics. Some references [**Tao07, TV06**] were already suggested in Section 3.2. A recent development is based on Gowers' idea to consider, in addition to "linear phase functions" $x \mapsto e(ax)$, also "higher-order phase functions" such as $x \mapsto e(ax^2)$. A nice account of this is given by Tao [**Tao12**].

In Section 6 we gave a glimpse of the theory of Boolean functions. The reader interested in diving deeper into this area should enjoy a survey by Kalai et al. [**KS06**], a recent monograph by O'Donell [**O'D14**], or online lecture notes by Hatami [**Hat14**].

Acknowledgment. We would like to thank Martin Böhm, Martin Mareš, and Vojta Tůma for reading and valuable comments.

Bibliography

[AB09] S. Arora and B. Barak. *Computational Complexity: A Modern Approach.* Cambridge University Press, Cambridge, 2009.

[Bis11] P. Biswal. Hypercontractivity and its applications. arXiv:1101.2913, 2011.

[BLR93] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. Comput. System Sci.*, 47(3):549–595, 1993.

[Gol08] O. Goldreich. *Computational Complexity: A Conceptual Perspective.* Cambridge University Press, Cambridge, 2008.

[GW10] B. Green and J. Wolf. A note on Elkin's improvement of Behrend's construction. In *Additive Number Theory*, pages 141–144. Springer, New York, 2010.

[Hat14] H. Hatami. Harmonic analysis of Boolean functions, 2014. Available at the author's web page `http://cs.mcgill.ca/~hatami/comp760-2014/lectures.pdf`.

[KKL88] J. Kahn, G. Kalai, and N. Linial. The influence of variables on Boolean functions (Extended abstract). In *29th Annual Symposium on Foundations of Computer Science*, pages 68–80, 1988.

[Knu73] D. E. Knuth. *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms*. Addison-Wesley, 1973.

[Kör89] T. W. Körner. *Fourier Analysis*. Cambridge University Press, Cambridge, second edition, 1989.

[KS06] G. Kalai and S. Safra. Threshold phenomena and influence: perspectives from mathematics, computer science, and economics. In *Computational Complexity and Statistical Physics*, St. Fe Inst. Stud. Sci. Complex., pages 25–60. Oxford Univ. Press, New York, 2006.

[Mes95] R. Meshulam. On subsets of finite abelian groups with no 3-term arithmetic progressions. *J. Combin. Theory Ser. A*, 71(1):168–172, 1995.

[MG02] D. Micciancio and S. Goldwasser. *Complexity of Lattice Problems: A Cryptographic Perspective*. The Kluwer International Series in Engineering and Computer Science, 671. Kluwer Academic Publishers, Boston, MA, 2002.

[O'D14] R. O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, Cambridge, 2014.

[Tao07] T. Tao. The ergodic and combinatorial approaches to Szemerédi's theorem. In *Additive Combinatorics*, volume 43 of *CRM Proc. Lecture Notes*, pages 145–193. Amer. Math. Soc., Providence, RI, 2007.

[Tao12] T. Tao. *Higher Order Fourier Analysis*, volume 142 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

[Tao14] T. Tao. 245C, Notes 2: The Fourier transform, 2014. Available at the author's blog `https://terrytao.wordpress.com/2009/04/06/the-fourier-transform/`.

[Ter99] A. Terras. *Fourier Analysis on Finite Groups and Applications*, volume 43 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1999.

[TV06] T. Tao and V. Vu. *Additive Combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.

Chapter 4

Representations of Finite Groups

We assume that the reader is somewhat familiar with groups, but we will hardly use anything from group theory except for the definition of a group and several examples.

The idea of group representations is to assign an invertible matrix to every element of a group G in such a way that the group operation corresponds to matrix multiplication. In this way, problems about G become problems about matrices, where one can often apply powerful tools from linear algebra. Moreover, for all finite groups and many others, every representation can be decomposed into basic building blocks called *irreducible representations*. Knowing all possible irreducible representations for G allows one to express every complex function defined on G in a "nice" basis. In this way, for abelian (i.e., commutative) groups G we recover Fourier analysis on G, and for noncommutative groups we obtain a "noncommutative Fourier analysis" on G. The usual, commutative Fourier analysis that is the theme of Chapter 3, is a special case: one where we consider representations of abelian groups by 1×1 complex matrices.

Applications of group representations in computer science and combinatorics are not (yet) too numerous. But they are usually beautiful and often yield results for which no alternative approach

is known. Selected applications will be mentioned in Section 6 below, together with textbooks on representation theory.

Representations can be studied not only for groups, but also for other, more general, algebraic structures. For a deeper treatment of representation theory with this added generality we refer to the book of Etingof et al. [**EGH⁺11**].

1. Basic Definitions and Examples

The notion of representation. Let V be a vector space over field $\mathbb{K}$. For now $\mathbb{K}$ is arbitrary, but later on we will work only with $\mathbb{K} = \mathbb{C}$, the complex numbers. We will assume throughout that V is nontrivial; i.e., $V \neq \{0\}$.

Let $\mathrm{GL}(V)$ denote the set of all invertible linear maps $V \to V$. With the operation of composition of maps, $\mathrm{GL}(V)$ is a group, the *general linear group* of V.

A **representation** of a group G is a homomorphism

$$\rho\colon G \to \mathrm{GL}(V)$$

for some vector space V.

In words, a representation ρ sends every element $g \in G$ to an invertible linear map $\rho(g)\colon V \to V$, in such a way that $\rho(gh) = \rho(g)\rho(h)$ for every $g, h \in G$, where $\rho(g)\rho(h)$ denotes the composition of $\rho(g)$ and $\rho(h)$. It follows that $\rho(g^{-1}) = \rho(g)^{-1}$ and $\rho(e)$ is the identity map, where e is the unit element of G. (In a still somewhat different language, we can also say that a representation of G is an action of G on a vector space by linear maps.)

For aesthetic reasons, we will usually write ρ_g instead of $\rho(g)$.

Standing assumptions: throughout this chapter we assume that the group G is *finite* and that V is a *finite-dimensional* nontrivial vector space over the field $\mathbb{C}$ of *complex numbers.*

With these assumptions, the basic theory comes out relatively simple. Most of the results below hold in greater generality, though, as can be deduced from the proofs or found in textbooks.

If we fix a basis B of V, every $f \in \mathrm{GL}(V)$ can be represented by a square matrix, which is invertible (i.e., nonsingular). The choice of B thus induces an isomorphism $\mathrm{GL}(V) \cong \mathrm{GL}(\mathbb{C}, \dim V)$, where $\mathrm{GL}(\mathbb{C}, n)$ is the group of all complex invertible $n \times n$ matrices. So we can also view a representation as an assignment of invertible matrices to group elements as announced above.

Although in theory we deal with complex matrices, in examples we will see mostly real ones. This is similar to introductions to eigenvalues of matrices, where one usually works in the complex setting but most of the concrete examples are real.

The **dimension** of a representation $\rho\colon G \to \mathrm{GL}(V)$ is defined as $\dim V$. If the mapping ρ is clear from the context, we sometimes talk about the space V itself as the representation of G.

***G*-linear maps and equivalence.** As usual in mathematics, having defined a class of objects, in our case representations of G, we also want to introduce a notion of "good" mappings (morphisms) between these objects.

Here they are called *G-linear maps*. If V, W are vector spaces (over the same field), $\rho\colon G \to \mathrm{GL}(V)$ and $\tau\colon G \to \mathrm{GL}(W)$ are representations of a group G, and $f\colon V \to W$ is a linear map, we say that f is ***G*-linear** if $f\rho_g = \tau_g f$ for every $g \in G$. Many mathematicians would probably prefer to express this condition by drawing the following diagram

$$\begin{array}{ccc} V & \xrightarrow{\rho_g} & V \\ \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle f} \\ W & \xrightarrow{\tau_g} & W \end{array}$$

and saying that it *commutes* for every $g \in G$.

Two representations ρ and τ of G are **equivalent** if there is an isomorphism of the two vector spaces that preserves the action of G. That is, there exists a G-linear isomorphism between V and W.

Basic examples. The simplest example of a representation of any group G is the **trivial representation** that assigns the identity map to every element $g \in G$. This illustrates that representations need not be injective.

Next, let S_n denote the **symmetric group** of all permutations of the set $\{1, 2, \ldots, n\}$ with composition as the group operation.

Let V be a vector space. The **alternating representation** of S_n sends each permutation $\pi \in S_n$ to the linear map $\tau_\pi \colon V \to V$ given by $\tau_\pi(v) = \mathrm{sgn}(\pi) \cdot v$. For $n \geq 2$, this is different from the trivial representation.

Here is another representation of S_n: Fix an n-dimensional vector space V and a basis $(e_1, \ldots, e_n)$ of V. The image of a permutation $\pi \in S_n$ is the mapping $\gamma_\pi \colon V \to V$ that permutes the basis elements according to π. That is, we set $\gamma_\pi(e_i) = e_{\pi(i)}$ and extend linearly to the rest of V. This is called the **permutation representation** of S_n.

For an arbitrary n-element group G, the so-called **regular representation** ρ is constructed as follows. Let V be an n-dimensional vector space with a basis $(e_g : g \in G)$ whose elements are indexed by the elements of G. Then, for each $h \in G$, the linear mapping $\rho_h \colon V \to V$ satisfies $\rho_h(e_g) = e_{hg}$ for every $g \in G$ (this condition and linearity determine f uniquely, of course). That is, ρ_h permutes the basis elements in the same way as the left multiplication by h permutes the elements of G.

This is in a way similar to the permutation representation of S_n mentioned before, but note that while the permutation representation of S_n has dimension n, the regular representation of S_n has dimension $|S_n| = n!$.

Remark: the group algebra. In more modern treatments, a group representation is usually viewed as a representation of the **group algebra** $\mathbb{C}[G]$. This is a simple but useful concept: $\mathbb{C}[G]$ is a complex vector space with the elements of G forming a basis (or, if you prefer, with basis vectors indexed by G), and with an operation of multiplication that is a linear extension of the operation on G.

Thus, $(\sum_{g\in G}\alpha_g g)(\sum_{h\in G}\beta_h h) = \sum_{g,h\in G}\alpha_g\beta_h(gh)$. Given a representation $\rho\colon G \to \mathrm{GL}(V)$ of G, we can extend it (linearly) to a map $\mathbb{C}[G] \to \operatorname{End} V$, where $\operatorname{End} V$ is the space of all linear maps $V \to V$ (including non-invertible ones), and this is the representation of $\mathbb{C}[G]$ alluded to above. All of this can be done for any field in place of $\mathbb{C}$, of course. Here we will stick to group representations as defined earlier.

2. Decompositions into Irreducible Representations

Let $\rho\colon G \to \mathrm{GL}(V)$ be a representation of G. A subspace $W \subseteq V$ is **invariant** with respect to ρ if $\rho_g(W) \subseteq W$ for every $g \in G$. The restriction of ρ to an invariant subspace W yields a representation $\rho_W : G \to \mathrm{GL}(W)$, a *subrepresentation* of ρ.

For example, let $(e_1, \ldots, e_n)$ be the standard basis of $\mathbb{C}^n$ and let $\rho\colon S_n \to \mathrm{GL}(\mathbb{C}^n)$ be the permutation representation of the symmetric group considered above. Thus, ρ_π acts by permuting the coordinates of a vector $v = (v_1, \ldots, v_n)$ according to π; i.e., $\rho_\pi(v) = (v_{\pi(1)}, \ldots, v_{\pi(n)})$. Then the 1-dimensional subspace $U = \{(t, t, \ldots, t) : t \in \mathbb{C}\}$ is invariant, and so is the subspace $W = \{v \in \mathbb{C}^n : \sum v_i = 0\}$, the orthogonal complement of U.

A representation $\rho\colon G \to \mathrm{GL}(V)$ is called **irreducible** if it has no proper subrepresentations; that is, if V has no nontrivial invariant subspaces (that is, invariant subspaces different from $\{0\}$ and V).

Exercise 2.1. For $n = 3$, prove that the restriction of the permutation representation to the subspace W defined above is irreducible. (This is true in general but may not be so easy without appropriate tools.) ✠

Irreducible representations play a crucial role in the theory of representations. They have two properties that make them extremely convenient to use—first, there is a lot of useful theory developed around them, and second, we can reduce problems involving general representations to problems involving irreducible ones.

Direct sums. We will now prove that every representation is a direct sum of irreducible representations. Later on, in Corollary 3.9, we will also see that this decomposition is unique in a suitable sense. So the role of irreducible representations is somewhat similar to the role of prime numbers in number theory.

First, we recall that the **direct sum** $V \oplus W$ of vector spaces V and W can be defined as the vector space whose vectors are the pairs (v, w), $v \in V$, $w \in W$, with coordinate-wise addition of vectors and multiplication by scalars.

For example, the real plane $\mathbb{R}^2$ can be obtained as the direct sum of two copies of $\mathbb{R}$.

The **direct sum** of two representations $\rho\colon G \to \mathrm{GL}(V)$ and $\tau\colon G \to \mathrm{GL}(W)$ of the same group is defined in the expected way: it is a mapping $\rho \oplus \tau\colon G \to V \oplus W$ defined componentwise, i.e.,

$$(\rho \oplus \tau)_g(v, w) = (\rho_g(v), \tau_g(w)).$$

We have $\dim(\rho \oplus \tau) = \dim \rho + \dim \tau$.

A formalistic remark. If V is a vector space and U, W are subspaces such that $U \cap W = \{0\}$ and $U \cup W$ spans V, then we also write $V = U \oplus W$, although strictly speaking, according to our definition of direct sum, $U \oplus W$ is only isomorphic to V, where the isomorphism sends (u, v) to $u + v$.

Similarly, if $\rho\colon G \to \mathrm{GL}(V)$ is a representation, U, W are invariant subspaces of V with $V = U \oplus W$, and ρ_U, ρ_W are the corresponding subrepresentations, we write $\rho = \rho_U \oplus \rho_W$, although strictly speaking, $\rho_U \oplus \rho_W$ is only equivalent to ρ.

Maschke's theorem. The key result for decomposing representations into irreducible ones is the following.

Theorem 2.2 (Maschke's theorem)**.** *Let $\rho\colon G \to \mathrm{GL}(V)$ be a representation, and let U be an invariant subspace of V. Then there is another invariant subspace W such that $V = U \oplus W$.*

To be on the safe side, let us remember that we assume G finite and V finite-dimensional and complex. However, a similar result holds

for many important infinite groups as well, and the field $\mathbb{C}$ can be replaced by any field of characteristic zero.

A straightforward induction using Maschke's theorem and the finite dimensionality of V yields the desired result.

Every representation can be written as a direct sum of irreducible representations.

Theorem 2.2 is a powerful result with a simple proof.

Proof of Theorem 2.2. Since V is a finite-dimensional complex vector space, it can be equipped with an inner product, which we denote by $\langle .,. \rangle_0$. Having an inner product, we can form the orthogonal complement[1] W_0 of U, and as is well known, we have $V = U \oplus W_0$.

Unfortunately, W_0 need not be invariant with respect to ρ in general. Thus, using ρ, we construct a better inner product $\langle .,. \rangle$, given by

$$\langle u, v \rangle := \sum_{g \in G} \langle \rho_g(u), \rho_g(v) \rangle_0.$$

(The axioms of inner product need to be checked, but this is routine and we leave it to the interested reader.)

This new inner product is obviously invariant with respect to ρ in the sense that $\langle \rho_g(u), \rho_g(v) \rangle = \langle u, v \rangle$. From this it is easily seen that if we define W as the orthogonal complement of U under $\langle .,. \rangle$, then W is invariant. $\square$

Exercise 2.3. The goal is to prove Maschke's theorem in a slightly different way, avoiding the inner product and yielding the result over every field of characteristic zero.

(a) What is wrong with the following proof sketch? We can always find a direct complement W_0 of U, i.e., a subspace W_0 with $V = U \oplus W_0$. (To see this, take a basis B_U of U, extend it to a basis B_V of V, and let W_0 be the span of $B_V \setminus B_U$.) In general, W_0 need not be invariant, so we make it invariant by averaging: $W := \{|G|^{-1} \sum_{g \in G} \rho_g(w) : w \in W_0\}$.

[1]The *orthogonal complement* of a subspace U of a vector space V is the set $\{w \in V : \langle u, w \rangle = 0 \text{ for all } u \in U\}$.

(b) Find a right way of averaging, to obtain a correct proof. Hint: take a projection $p\colon V \to U$, make it invariant by a suitable averaging over G (the first way of doing the averaging coming to mind may not work!), and define W as the kernel of the resulting map. ✠

Another example. We have already seen one example of a decomposition: the decomposition of the permutation representation of S_3 into $U \oplus W$, with U of dimension 1 and W of dimension 2. Here is another example.

Consider the alternating representation ρ of S_n on $\mathbb{C}^2$ given by $\rho_\pi(x, y) = \operatorname{sgn}(\pi)(x, y)$. The line $\{(x, x) : x \in \mathbb{C}\}$ is an invariant subspace. The orthogonal complement under the standard inner product, the line $\{(x, -x) : x \in \mathbb{C}\}$, is also invariant. This choice is not unique; for example, the line $\{(x, -2x) : x \in \mathbb{C}\}$ is another invariant complement.

Weyl's unitarity trick. We recall that a square matrix A is **unitary** if $A^{-1} = \overline{A}^T$ (the transposition of the complex conjugate). A unitary matrix is a complex analog of an orthogonal matrix, since its columns form an orthonormal system in $\mathbb{C}^n$.

The construction of the invariant inner product $\langle ., . \rangle$ in the above proof of Maschke's theorem has the following useful consequence:

Corollary 2.4 (Weyl's unitarity)**.** *Every (complex, finite-dimensional) representation is equivalent to a representation by unitary matrices.*

Proof. Let us consider a representation $\rho\colon G \to \mathrm{GL}(V)$, and let $\langle ., . \rangle$ be the invariant inner product mentioned above. Choose an orthonormal basis $B = (e_1, \ldots, e_n)$ of V (by the Gram–Schmidt orthogonalization).

We recall that if $u = a_1 e_1 + \cdots + a_n e_n$ and $v = b_1 e_1 + \cdots + b_n e_n$ are expressions of two vectors of V in an orthonormal basis B, then $\langle u, v \rangle = \sum_{i=1}^n a_i \overline{b_i}$; that is, the inner product is computed as the standard inner product of the coordinate vectors.

Let us fix $g \in G$ and let A be the matrix of ρ_g with respect to B. Thus, A has the coordinates of $\rho_g(e_i)$ in the i-th column. To verify

that A is unitary, we need to check that its columns are orthonormal (with respect to the standard inner product).

By the above, the standard inner product of the i-th and j-th columns equals $\langle \rho_g(e_i), \rho_g(e_j)\rangle$. By the invariance of $\langle .,.\rangle$, this equals $\langle e_i, e_j\rangle = \delta_{ij}$ (the Kronecker delta). □

An optional digression: the tensor product. A useful way of constructing new representations from old ones is the *tensor product* (but we will not really use it in the rest of this chapter). In particular, irreducible subrepresentations of tensor products are often interesting and not so easy to describe explicitly.

Given two finite-dimensional vector spaces V, W, their **tensor product** $V \otimes W$ can be described as follows. Choose a basis $(v_1, \ldots, v_k)$ of V and a basis $(w_1, \ldots, w_\ell)$ of W. Then $V \otimes W$ is the vector space of all formal linear combinations

$$\sum_{i=1}^{k} \sum_{j=1}^{\ell} c_{ij}(v_i \otimes w_j),$$

where the c_{ij} are coefficients (complex numbers in our setting) and the $v_i \otimes w_j$ are understood as purely formal symbols.

Thus, $V \otimes W$ is isomorphic to the vector space of all $k \times \ell$ matrices, and it has dimension $k\ell$.

Let us remark that $V \otimes W$ can also be defined in a more conceptual way, without first selecting bases in V and W, but that definition may perhaps look more frightening at first encounter, so we avoid it here.

So far we have considered $v_i \otimes w_j$ as a formal symbol. Now we want to define, for every $v \in V$ and $w \in W$, an element $v \otimes w \in V \otimes W$. To this end, we express v and w in the chosen bases, $v = \sum_{i=1}^{k} a_i v_i$ and $w = \sum_{j=1}^{\ell} b_j w_j$, and we set

$$v \otimes w := \sum_{i=1}^{k} \sum_{j=1}^{\ell} a_i b_j (v_i \otimes w_j).$$

(It may be worth remarking that not every element of $V \otimes W$ is expressible as $v \otimes w$ for some v and w.)

Now we can define the **tensor product** $\rho \otimes \tau$ of representations $\rho\colon G \to \mathrm{GL}(V)$ and $\tau\colon G \to \mathrm{GL}(W)$: the linear mapping $(\rho\otimes\tau)_g\colon V\otimes W \to V \otimes W$ sends $v_i \otimes w_j$ to $\rho_g(v_i) \otimes \tau_g(w_j)$.

Exercise 2.5. In order to familiarize yourself with the notions, check that this indeed defines a representation. ✠

3. Irreducible Decompositions, Characters, Orthogonality

Towards noncommutative Fourier analysis. We recall that if G is a finite abelian group, a *character* of G is simply a group homomorphism from G to $\mathbb{C}\setminus\{0\}$, the multiplicative group of complex numbers. All characters form an orthonormal basis of the vector space of all functions $\varphi\colon G \to \mathbb{C}$ with the inner product

$$\langle \varphi, \psi \rangle := \frac{1}{|G|} \sum_{g\in G} \varphi(g)\overline{\psi(g)}. \tag{28}$$

The *Fourier transform* of a function $\varphi\colon G \to \mathbb{C}$ is then the expression of φ as a linear combination of characters; the coefficient of a character χ in this expression is often denoted by $\widehat{\varphi}(\chi)$ and it can be computed by the Fourier transform formula $\widehat{\varphi}(\chi) = \langle \varphi, \chi \rangle$. There are several basic and very useful theorems about the Fourier transform, such as an inversion formula, a formula for convolution, Plancherel's formula (saying that the Fourier transform preserves the inner product), etc.—see Chapter 3.

We have promised a noncommutative analog of these things, and here is a preview of what to expect; for details, see Section 3.4.

A noncommutative (finite) group G also has characters: a character χ is again a complex function on G (but not necessarily a homomorphism), and this time each character is associated to a representation of G. Only characters of *irreducible* representations are really used in the noncommutative Fourier analysis, and so it is important to understand enough about all possible irreducible representations of G. Conveniently, two irreducible representations have the same character iff they are equivalent.

Characters of mutually nonequivalent irreducible representations again form an orthonormal system with respect to the inner product in Equation (28). However, there are not enough of them to form a basis of all complex functions on G. Indeed, we recall that two elements $g_1, g_2 \in G$ are **conjugate** if $g_2 = hg_1h^{-1}$ for some $h \in G$, and that conjugacy is an equivalence on G. It turns out that each character is constant on each conjugacy class of G, and the characters form a basis for the space of all *class functions*, i.e., ones that are constant on each conjugacy class; see the proof of Theorem 3.17 below.

For the space of all functions $G \to \mathbb{C}$, the Fourier basis is going to be different and more complicated; roughly speaking, it will be derived by taking all mutually nonequivalent irreducible representations ρ of G, and for each of them, representing the ρ_g by unitary matrices according to Weyl's unitarity (Corollary 2.4).

The Fourier transform of an arbitrary function $\varphi\colon G \to \mathbb{C}$ is then conveniently regarded as a *matrix-valued* function (or, more formally, the values are linear maps). Namely, the Fourier transform $\widehat{\varphi}$ assigns to each irreducible representation $\rho\colon G \to \mathrm{GL}(V)$ the following linear map $V \to V$ (not necessarily invertible):

$$\widehat{\varphi}(\rho) := \sum_{g \in G} \varphi(g)\rho_g.$$

Then there are a suitable inversion formula, convolution formula, and Plancherel formula; see the end of this section or [**Dia88**, Chap. 2].

Here we are not going to build all of this theory, but we will lay solid foundations for it.

3.1. Characters and Schur's Lemma. We recall that the **trace** $\mathrm{Tr}\, A$ of a square matrix A is the sum of the elements on the main diagonal of A. The trace $\mathrm{Tr}\, f$ of a linear map (endomorphism) $f\colon V \to V$ is defined as $\mathrm{Tr}\, A$, where A is the matrix of f with respect to some basis of V.

To see that this definition of $\mathrm{Tr}\, f$ is independent of the choice of the basis of V, we recall that if B is the matrix of f with respect to

another basis of V, then we have $A = TBT^{-1}$ for some nonsingular matrix T. Since, as is not hard to check, $\mathrm{Tr}(XY) = \mathrm{Tr}(YX)$ for square matrices X and Y, we have $\mathrm{Tr}\, A = \mathrm{Tr}(TBT^{-1}) = \mathrm{Tr}(T^{-1}TB) = \mathrm{Tr}\, B$.

Characters. Let $\rho\colon G \to \mathrm{GL}(V)$ be a representation.

The **character** of ρ is the function $\chi\colon G \to \mathbb{C}$ given by

$$\chi(g) := \mathrm{Tr}\,\rho_g.$$

Here are some simple properties of characters; in particular we will see that characters are, indeed, class functions.

Lemma 3.1. *Let χ be a character of a d-dimensional representation.*

(i) $\chi(e) = d$, *where e is the unit element of G.*

(ii) $\chi(g^{-1}) = \overline{\chi(g)}$.

(iii) *If g_1, g_2 are conjugate in G, then $\chi(g_1) = \chi(g_2)$.*

Sketch of proof. Part (i) is obvious since $\rho_e = \mathrm{id}_V$ and its matrix is I_d, the identity matrix. Part (ii) can be easily verified using Weyl's unitarity (Corollary 2.4), and (iii) follows from the already mentioned formula $\mathrm{Tr}(TBT^{-1}) = \mathrm{Tr}(B)$. □

Exercise 3.2. Verify the following formulas for the characters of direct sum and tensor product: $\chi(\rho \oplus \tau) = \chi(\rho) + \chi(\tau)$ and $\chi(\rho \otimes \tau) = \chi(\rho) \cdot \chi(\tau)$. ✠

Next, we would like to show that the characters of nonequivalent irreducible representations are orthogonal. For this, though, we first need to build some tools.

Schur's lemma. This is a key device for understanding irreducible representations.

Theorem 3.3 (Schur's lemma)**.** *Let $\rho\colon G \to \mathrm{GL}(V)$ and $\tau\colon G \to \mathrm{GL}(W)$ be two* irreducible *representations and let $f\colon V \to W$ be a G-linear mapping (with respect to ρ and τ).*

(1) *Either f is an isomorphism (i.e., ρ and τ are equivalent), or $f = 0$.*

(2) *If $V = W$ and $\rho = \tau$, then f is a constant times the identity.*

Part (2) can also be rephrased as follows: *If $\rho\colon G \to \mathrm{GL}(V)$ is an irreducible representation and $f\colon V \to V$ is a linear map that commutes with all the ρ_g, then $f = c \cdot \mathrm{id}_V$.*

Proof. Since ρ is irreducible and $\ker f$ is an invariant subspace of V, we have either $\ker f = V$, which means that $f = 0$, or $\ker f = \{0\}$, and then f is injective. Similarly, invariant subspace $f(V)$ has to be either $\{0\}$ or W, meaning that $f = 0$ or f is surjective. It follows that $f = 0$ or f is an isomorphism.

As for (2), we note that since we are working over $\mathbb{C}$, every linear mapping on a nontrivial vector space has an eigenvalue (this is the main reason for preferring representations over $\mathbb{C}$). Let λ be an eigenvalue of f; this means that the mapping $f_1 = f - \lambda\,\mathrm{id}_V$ has a nontrivial kernel. Since f_1 is G-linear, $\ker f_1$ is an invariant subspace of V, and hence $\ker f_1 = V$ by irreducibility. So $f_1 = 0$ and $f = \lambda\,\mathrm{id}_V$. □

Here is an interesting corollary, which can be used to show that, for G abelian, the "noncommutative Fourier analysis" outlined above actually reduces to the usual Fourier analysis.

Corollary 3.4. *Every irreducible complex representation ρ of a finite abelian group G is one-dimensional.*

Proof. Since G is commutative, we have $\rho_g\rho_h = \rho_h\rho_g$ for every $g, h \in G$, and in particular, each ρ_g is a G-linear mapping.

So by part (2) of Schur's lemma we have $\rho_g(v) = c_g v$ for all $v \in V$ and some $c_g \in \mathbb{C}$. Fixing a nonzero $v \in V$, we get that

$W := \{cv : c \in \mathbb{C}\}$ is a 1-dimensional invariant subspace. Since ρ is irreducible, we have $V = W$. □

Exercise 3.5. In Fourier analysis, the basic notion is that of a character of an abelian group G: a mapping to the group $\mathbb{T}$ (unit complex numbers with multiplication) that is a group homomorphism. Explain which object in the proof of the previous corollary corresponds to this definition of a character. ✠

3.2. Orthogonality. Here we prove certain orthogonality relations for matrices of irreducible representations, and deduce orthogonality of characters from them.

Let us introduce a notational convention: if $\rho\colon G \to \mathrm{GL}(V)$ is an irreducible representation and we fix some basis B of V, we write R_g for the matrix of ρ_g with respect to B, and $r_{ij}(g)$ stands for the element of R_g at position (i, j). Similarly, for a representation τ we use T_g and $t_{ij}(g)$.

For fixed i and j, we can regard r_{ij} as a function $G \to \mathbb{C}$. For such functions, we use the inner product defined by Eq. (28) above. We will find that, with a suitable choice of bases, they are orthogonal, as is described in the next proposition.

Proposition 3.6. *Let $\rho\colon G \to \mathrm{GL}(V)$ and $\tau\colon G \to \mathrm{GL}(W)$ be two irreducible representations, and let us choose, according to Weyl's unitarity (Corollary 2.4), bases of V and W so that all of the matrices R_g and T_g are unitary. Then:*

(1) *If ρ and τ are nonequivalent, then $\langle r_{ij}, t_{k\ell}\rangle = 0$, for all i, j, k, ℓ.*

(2) *We have*

$$\langle r_{ij}, r_{k\ell}\rangle = \begin{cases} \frac{1}{d} & \text{if } (i,j) = (k,\ell) \\ 0 & \text{otherwise,} \end{cases}$$

where $d = \dim V$.

Thus, we can see that the functions r_{ij} obtained from various nonequivalent irreducible representations form an orthogonal system.

Later, we will see that if we take *all* nonequivalent irreducible representations, we obtain an orthogonal basis of the space of all functions $G \to \mathbb{C}$; this is the construction mentioned at the beginning of this section.

Proof. Let $f\colon V \to W$ be a linear mapping. Let us form the following averaged version of f:

$$\tilde{f} = \frac{1}{|G|} \sum_{g\in G} \tau_g^{-1} f \rho_g.$$

It is straightforward to verify that $\tilde{f}$ is G-linear (by the way, this is the right way of averaging for Exercise 2.3(b)).

If ρ and τ are not equivalent, then $\tilde{f} = 0$ by part (1) of Schur's lemma.

If $V = W$ and $\rho = \tau$ with dimension d, then Schur's lemma yields $\tilde{f} = c \cdot \mathrm{id}_V$. To calculate c, we calculate traces: from $\tilde{f} = c \cdot \mathrm{id}_V$ we get $\operatorname{Tr} \tilde{f} = cd$, while from the definition of $\tilde{f}$ we obtain $\operatorname{Tr} \tilde{f} = (1/|G|) \sum_g \operatorname{Tr}(\rho_g^{-1} f \rho_g) = (1/|G|) \sum_g \operatorname{Tr} f = \operatorname{Tr} f$. Hence

$$\tilde{f} = \frac{\operatorname{Tr} f}{d} \,\mathrm{id}_V \,. \tag{29}$$

Now let R_g, T_g be the matrices as in the notational conventions above, and let F and $\tilde{F}$ be the matrices of f and $\tilde{f}$, respectively, all with respect to the chosen bases. Rewriting $\tilde{f} = 0$ in matrix form yields, for every i, j,

$$0 = \tilde{f}_{ij} = \frac{1}{|G|} \sum_{g\in G} \sum_{p,q=1}^{n} t_{ip}(g^{-1}) \cdot f_{pq} \cdot r_{qj}(g).$$

We are free to choose f in any way we like, so we choose indices k, ℓ arbitrarily and set $f_{k\ell} = 1$ and $f_{i,j} = 0$ whenever $(i, j) \neq (k\ell)$. Then we get

$$\frac{1}{|G|} \sum_{g\in G} t_{ik}(g^{-1}) r_{\ell j}(g) = 0 \text{ for every } i, j, k, \ell. \tag{30}$$

Now if we assume T_g unitary, we have $t_{ik}(g^{-1}) = \overline{t_{ki}(g)}$, and then (30) becomes $\langle r_{\ell j}, t_{ki} \rangle = 0$; this is the relation we wanted, only the names of the indices have not come out quite as intended.

The calculation of $\langle r_{ij}, r_{k\ell}\rangle$ is done in the same way starting with (29), and we leave this to the reader. □

Orthogonality of characters. The promised orthogonality of characters follows immediately from Proposition 3.6.

Corollary 3.7. *The characters of nonequivalent irreducible representations are orthonormal.*

Proof. Let χ and χ' be the characters of irreducible representations ρ and ρ', respectively. Choose bases so that the matrices R_g and R'_g are unitary. Then $\chi(g) = \operatorname{Tr} R_g = \sum_{i=1}^{d} r_{ii}(g)$, and similarly for χ'. So using Proposition 3.6,

$$\langle \chi, \chi' \rangle = \sum_{i,j=1}^{d} \langle r_{ii}, r'_{jj} \rangle = \sum_{i=1}^{d} \langle r_{ii}, r'_{ii} \rangle.$$

All terms in the last sum are 0 for ρ and ρ' nonequivalent, and they equal $\frac{1}{d}$ for $\rho = \rho'$. □

3.3. Characters and Decomposition into Irreducible Representations.

Proposition 3.8. *Let $\rho\colon G \to \mathrm{GL}(V)$ be a representation with character χ. Let $V = W_1 \oplus \cdots \oplus W_k$ be a decomposition of V into irreducible representations. If W is an irreducible representation with character χ_i, then the number of the W_i equivalent to W equals $\langle \chi, \chi_i \rangle$.*

Proof. We have $\chi = \chi_1 + \cdots + \chi_k$ by Exercise 3.2, and so $\langle \chi, \chi_i \rangle = \sum_{i=1}^{k} \langle \chi_i, \chi_i \rangle$. By the orthonormality of characters, each of the terms is either 0 or 1, depending on whether W_i is equivalent to W. □

Corollary 3.9.

(i) (Uniqueness of decomposition) *The decomposition of a representation* $\rho\colon G \to V$ *into irreducible representations is unique, in the sense that if* $V = V_1 \oplus \cdots \oplus V_k$ *and* $V = W_1 \oplus \cdots \oplus W_\ell$ *are two such decompositions, then every irreducible representation occurs the same number of times in both.*

(ii) (Character determines representation) *Two representations are equivalent iff they have the same character.*

(iii) (Irreducibility criterion) *A representation is irreducible iff its character* χ *satisfies* $\langle \chi, \chi \rangle = 1$.

Proof. Part (i) follows immediately from Proposition 3.8 since $\langle \chi, \chi_i \rangle$ does not depend on the irreducible decomposition. So does (ii), since two representations with the same character have the same irreducible decomposition. As for (iii), if $V = m_1 W_1 \oplus \cdots \oplus m_k W_k$, where the W_i are nonequivalent irreducible representations,[2] we have $\langle \chi, \chi \rangle = \sum_{i=1}^k m_i^2$. This equals 1 if and only if ρ is irreducible. □

Exercise 3.10. Now do Exercise 2.1 for n arbitrary. ✠

Decomposing the regular representation. We recall that in a regular representation, we choose a vector space V of dimension $|G|$ with basis $(e_g : g \in G)$, and we define $\rho_h(e_g) = e_{hg}$. Let $W_1, \ldots, W_k$ be a maximal set of nonequivalent irreducible representations of G. Denote by χ_i the character of W_i, and by d_i its dimension.

Lemma 3.11. *The character* r_G *of the regular representation is given by* $r_G(e) = |G|$ *and* $r_G(g) = 0$ *whenever* $g \neq e$.

Proof. The first statement is easy. For the second one, consider the matrix of ρ_g with respect to the basis $(e_g : g \in G)$; we claim that its diagonal is zero. If not, there is a $g \in G$ such that $\rho_h(e_g) = e_g$, but this means that $hg = h$, and so $g = e$, which is a contradiction. □

Corollary 3.12. *Every irreducible representation is contained in the regular representation with multiplicity equal to its dimension.*

[2] By $3W$ we mean $W \oplus W \oplus W$.

Proof. $\langle r_g, \chi_i \rangle = d_i$. □

Corollary 3.13. *$\sum d_i^2 = |G|$, and whenever $g \neq e$, $\sum d_i \chi_i(g) = 0$.*

Proof. According to the previous corollary, $r_G(g) = \sum d_i \chi_i(g)$. Substituting $g = e$ and $g \neq e$, we get the two statements. □

Let r_{ij} be the orthogonal functions obtained from unitary matrices of irreducible representations as in Proposition 3.6.

Corollary 3.14. *The functions r_{ij} for all nonequivalent unitary irreducible representations form a basis for the vector space of all functions $G \to \mathbb{C}$.*

Proof. We already know that the functions are orthogonal, and hence also linearly independent. There are $\sum d_i^2 = |G|$ of them. This is equal to the dimension of the vector space in question. □

3.4. Noncommutative Fourier Analysis. Here we just state the main formulas without proofs, referring to [**Dia88**, Chap. 2].

We first recall the Fourier transform:

$$\widehat{\varphi}(\rho) = \sum_{g \in G} \varphi(g) \rho_g$$

(note that $\widehat{\varphi}(\rho)$ is a linear map $V \to V$).

Let $\rho^{(1)}, \dots, \rho^{(k)}$ be a list of all nonequivalent irreducible representations of G, and let $d_i = \dim \rho^{(i)}$. Then the Fourier inversion formula, the noncommutative version of Theorem 2.5 from Chapter 3, reads

$$\varphi(g) = \frac{1}{|G|} \sum_{i=1}^{k} d_i \operatorname{Tr}\left(\rho^{(i)}_{g^{-1}} \widehat{\varphi}(\rho^{(i)}) \right).$$

The *convolution* $\varphi * \psi$ of two functions on G is given by

$$(\varphi * \psi)(g) = \sum_{h \in G} \varphi(gh^{-1}) \psi(h),$$

the usual formula from the abelian case, actually. Textually, the convolution formula for the Fourier transform is also the same as in the abelian case:

$$\widehat{\varphi * \psi}(\rho) = \widehat{\varphi}(\rho) \widehat{\psi}(\rho).$$

Exercise 3.15. Prove that. ✠

With the tools built so far, checking these results is not too hard and the reader may try that as an exercise.

Exercise 3.16. Find all irreducible representations of the cyclic group C_n (use Corollary 3.4) and their characters. Check that these are the characters of C_n in the usual sense for abelian groups, and that the noncommutative Fourier transform formula coincides with the commutative one. ✠

Irreducible representations and conjugacy classes. Using the Fourier transform, one can obtain the following remarkable counting result.

Theorem 3.17. *The number of mutually nonequivalent irreducible representations of a finite group G equals the number of conjugacy classes of G.*

The proof is by a dimension-counting argument; interestingly, no explicit bijection between the conjugacy classes and the irreducible representations is known.

Proof. We again argue via the dimension of a suitable vector space.

The number of conjugacy classes equals the dimension of the vector space of all class functions $G \to \mathbb{C}$ (i.e., functions that are constant on each conjugacy class). We also know that the characters of nonequivalent irreducible representations are linearly independent class functions.

Thus, to finish the proof, it suffices to show that characters span all class functions. For technical reasons, we will actually prove that the *complex conjugates* $\overline{\chi}$ of all characters span all class functions. To do this, we start by assuming that f is a class function orthogonal to all $\overline{\chi}$, and we will verify that f must be identically 0.

To this end, we show that the Fourier transform $\widehat{f}$ is identically 0 and appeal to the inversion formula to conclude $f = 0$.

So let us fix an irreducible representation $\rho\colon G \to \mathrm{GL}(V)$ and write

$$F = \hat{f}(\rho) = \sum_{g \in G} f(g)\rho_g.$$

The next step is left to the reader:

Exercise 3.18. Using the assumption that f is a class function, i.e., $f(hgh^{-1}) = f(g)$ for all $g, h \in G$, show that F is a G-linear map, that is, $F\rho_h = \rho_h F$ for all $h \in G$. ✠

By Schur's lemma, we then have $F = c \cdot \mathrm{id}$, and it remains to show $c = 0$, which we do by calculating $\mathrm{Tr}\, F = \sum_{g\in G} f(g)\, \mathrm{Tr}\, \rho_g = \sum_{g\in G} f(g)\chi(g) = \langle f, \overline{\chi}\rangle = 0$ by the orthogonality assumption, where χ is the character of ρ. □

4. Irreducible Representations of the Symmetric Group

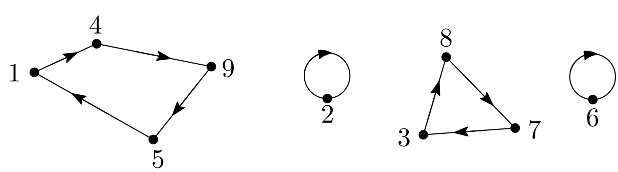
Representations of S_n are particularly useful, and they have been applied in many problems involving permutations. The application presented in the next section is also of this kind.

Representation theory of the symmetric group is rich and classical, and presenting the basics requires a nontrivial amount of work. We provide just a glimpse of the subject without proofs.

Conjugacy classes of permutations. First, we recall the cycle structure of a permutation. A permutation is a bijection

$$\pi\colon \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}.$$

If we draw dots numbered 1 through n and make an arrow from i to $\pi(i)$ for every i,



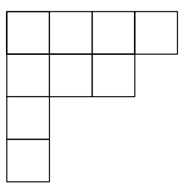
then every dot has one incoming and one outgoing arrow, and thus the picture consists of several disjoint cycles, the cycles of π. Let $\lambda(\pi) =$

$(\lambda_1, \ldots, \lambda_k)$ be the list of all cycle lengths written in nondecreasing order (so $\lambda_1 + \lambda_2 + \cdots + \lambda_k = n$); for the π depicted above $\lambda(\pi) = (4, 3, 1, 1)$.

Exercise 4.1. Show that two permutations $\pi, \pi' \in S_n$ are conjugate iff $\lambda(\pi) = \lambda(\pi')$.

This exercise shows that conjugacy classes of S_n are in one-to-one correspondence with partitions of n. This suggests that the irreducible representations of S_n should be related to partitions; at least we know, by Theorem 3.17, that the number of irreducible representations equals the number of partitions of n. And indeed, for every partition λ of n one can define an irreducible representation, called a *Specht module* and usually denoted by S^λ, and in this way one obtains all irreducible representations. To introduce S^λ we need several auxiliary notions.

Tableaus and tabloids. First we represent λ, a partition of n, by its **Ferrer's diagram**; for our $\lambda = (4, 3, 1, 1)$ it looks like this:



A **tableau** T is obtained by filling the squares of a Ferrer diagram with numbers 1 through n, using each exactly once, for example like this:

8	5	2	7
3	9	6	
1			
4			

So there are $n!$ tableaus for a given λ.

A **tabloid** is an equivalence class of tableaus, where two tableaus are equivalent if one can be obtained from the other by permuting

the numbers in each row. The tabloid t containing a tableau T is also denoted by $\{T\}$. We can represent a tabloid by the tableau where each row is in increasing order:

2	5	7	8
3	6	9	
1			
4			

Let $\mathrm{tbd}(\lambda)$ be the set of all tabloids for given λ; it is easy to see that their number is $\frac{n!}{\lambda_1!\cdots\lambda_k!} = \binom{n}{\lambda_1,\ldots,\lambda_k}$ (a multinomial coefficient).

If we take a tableau T and permute its entries according to a permutation π, we get a new tableau, which we denote by πT. This also extends to tabloids, $\pi\{T\} = \{\pi T\}$. So we can define a representation of S_n as follows: we take a vector space M^λ with a basis indexed by all tabloids, i.e., $(e_t : t \in \mathrm{tbd}(\lambda))$.

The action of S_n sending e_t to $e_{\pi t}$ turns M^λ into a representation of S_n, but this is not yet the irreducible representation S^λ we are after. The Specht module S^λ is an invariant subspace of M^λ spanned by the vectors to be defined next.

Let T be a tableau, and let $C_T \subseteq S_n$ be the set of those permutations that, when applied to T, only permutes the numbers in each *column*. For the first tableau drawn in this section, C_T consists of the permutations that permute $1,3,4,8$ among themselves, may exchange 5 and 9 as well as 2 and 6, and keep 7 fixed. For each tableau T, we define a vector $w_T \in M^\lambda$ as the following linear combination of the basis elements:

$$w_T := \sum_{\pi \in C_T} \mathrm{sgn}(\pi) \cdot e_{\{\pi T\}}.$$

The Specht module S^λ is the span of the w_T in M^λ, with the inherited action of S_n.

We note that the w_T are far from linearly independent. Actually, it is known that a basis of S^λ, the so-called *Specht basis*, is obtained by taking only those w_T for which T is a **standard tableau**, meaning

that the numbers are increasing in each row and each column, as in the next picture:

1	3	5	8
2	6	9	
4			
7			

So $\dim S^\lambda$ is the number of standard tableaus for λ, a well-studied combinatorial quantity.

For proofs that the S^λ are irreducible, mutually nonequivalent, and they exhaust all irreducible representations, we refer, e.g., to [**Dia88**, Chap. 7].

Exercise 4.2. Describe the Specht module S^λ and the action of S_n on it explicitly for $\lambda = (n)$, $\lambda = (1, 1, \ldots, 1)$, and $\lambda = (n-1, 1)$. Can some of them be identified with irreducible representations of S_n mentioned in the previous sections? ✠

5. An Application in Communication Complexity

The basic scenario in communication complexity is the following: A real-valued function f defined on $\{1, \ldots, N\}^2$ is known to both Alice and Bob. Alice is given a number x while Bob is given y, both integers between 1 and N. Their task is to determine together the value $f(x, y)$ while having communicated as little as possible.

Knowing f, but before they are given x and y, they decide on a communication protocol. This protocol specifies, in each step, which of the players sends information to the other one, and the information sent. The information sent by Alice depends on x and on the previous messages received from Bob, and similarly for Bob. The goal is that at the end of their communication, both players know the value $f(x, y)$. The communication complexity of the protocol is the maximum, over all choices of x and y, of the number of bits exchanged, and the **communication complexity** of f, denoted by $\mathrm{CC}(f)$, is the smallest possible communication complexity of a protocol.

More precisely, we speak of *deterministic* communication complexity, since the protocol is assumed to be deterministic. In the literature, *randomized* communication complexity is also studied, where Alice and Bob can use a random number generator and the answer needs to be correct only with probability sufficiently close to 1, *quantum* communication complexity, etc.

Two examples. Consider the function $f_1(x, y) = (x + y) \bmod 2$, i.e., the parity of $x + y$. For both players to know the answer, information has to flow in both directions, so $\mathrm{CC}(f_1) \geq 2$. On the other hand, it suffices to exchange two bits—each player sends the parity of his input.

As a second example, consider the function f_2 that equals 1 if $x = y$ and 0 otherwise (equality). It is not hard to see that at least $\log_2 N$ bits of communication are necessary.

The log-rank conjecture. Let us represent the function f by an $N \times N$ matrix $M = M(f)$ with $m_{ij} = f(i, j)$. The rank of M also measures the complexity of f in a certain sense, and it provides a powerful lower bound for the communication complexity. Namely, by a well known result of Mehlhorn and Schmidt [**MS82**],

$$\mathrm{CC}(f) \geq \log_2 \operatorname{rank} M$$

(where the rank can be taken over any field).

Lovász and Saks [**LS93**] asked whether $\mathrm{CC}(f)$ can also be bounded *from above* by $(\log_2 \operatorname{rank} M)^c + 2$ for some constant c (independent of n and f), where this time the rank is taken over the reals.[3] This has become known as the *log-rank conjecture*, and it is one of the most important open problems in communication complexity.

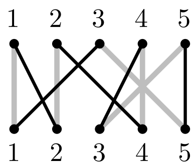
It was difficult to find f with $\mathrm{CC}(f)$ substantially bigger than $\log_2 \operatorname{rank} M$; indeed, the first gap found was by a multiplicative constant. The current record, at the time of writing, is $\mathrm{CC}(f)$ of order $(\log_2 \operatorname{rank} M)^C$ with $C \approx 1.63$, an improvement by Kushilevitz of a construction of Nisan and Wigderson; see [**NW95**].

Here we present a part of an earlier, weaker result of Raz and Spieker [**RS95**], which was the first to show that $\mathrm{CC}(f)$ cannot be

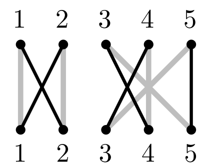
[3]The "+2" in the formula is needed to deal with the cases where $\operatorname{rank} M \leq 2$.

bounded by $O(\log_2 \operatorname{rank} M)$. Namely, they produced, for every n, an f with $\log_2 \operatorname{rank} M = O(n)$ and $\mathrm{CC}(f) = \Theta(n \log\log n)$.

Their f is described in the following way. Each of the two players is given a perfect matching in the complete bipartite graph $K_{n,n}$. The function is 1 if the union of the two matchings is a Hamiltonian cycle (i.e., a cycle passing through all vertices), and 0 otherwise.



$f = 1$



$f = 0$

A perfect matching in $K_{n,n}$ can be regarded as a permutation $\sigma \in S_n$ (the edges connect vertex i in the top level to vertex $\sigma(i)$ in the bottom level). Thus, we have $N = n!$, and the corresponding matrix M has rows and columns indexed by permutations of n.

Moreover, it is easy to see that the two perfect matchings corresponding to permutations σ and τ form a Hamiltonian cycle iff the composition $\sigma\tau^{-1}$ is a permutation with a single cycle.

We will use representation theory to compute $\operatorname{rank} M$:

Theorem 5.1. *The rank of M is $\binom{2n-2}{n-1} \leq 2^{2(n-1)}$.*

The second part of the Raz–Spieker result, estimating the communication complexity, is done by an information-theoretic argument and we do not consider it here.

To prove the theorem, we start by considering the regular representation ρ of S_n, on a vector space V of dimension $|S_n| = n!$, with basis $B = (e_\sigma : \sigma \in S_n)$ indexed by permutations. We recall that $\rho_\pi(e_\sigma) = e_{\pi\sigma}$.

Let R_π be the matrix of ρ_π with respect to the basis B, and let $X_n \subseteq S_n$ be the set of all unicyclic permutations.

We claim that

$$M = \sum_{\pi \in X_n} R_\pi,$$

where M is the matrix in the theorem. Checking this is just unwrapping the definitions: as we have seen, $M_{\sigma\tau} = 1$ iff $\sigma\tau^{-1} \in X_n$, while $(R_\pi)_{\sigma\tau} = 1$ means that $\pi\tau = \sigma$, i.e., $\pi = \sigma\tau^{-1}$.

Next, let $\mu\colon V \to V$ be the linear mapping whose matrix with respect to the basis B is M. Hence $\mu = \sum_{\pi\in X_n} \rho_\pi$.

Lemma 5.2. *The mapping μ is S_n-linear; in other words, $MR_\pi = R_\pi M$ for every $\pi \in S_n$.*

Exercise 5.3. Prove that. ✠

Now we let $\rho = \rho^{(1)} \oplus \cdots \oplus \rho^{(k)}$ be a decomposition of ρ into irreducible representations, let W_i be the invariant subspace corresponding to $\rho^{(i)}$, and let $d_i = \dim W_i$.

Since W_i is invariant under the ρ_π and μ is a linear combination of the ρ_π, we also have W_i invariant under μ. In particular, we can define $\mu^{(i)}$ as the restriction of μ to W_i.

The rank of M we are after equals $\dim \mu(V) = \sum_{i=1}^{k} \dim \mu^{(i)}(W_i)$. (If you like, think of a new basis whose first d_1 vectors span W_1, the next d_2 ones span W_2; etc. Then the matrix of μ with respect to this new basis is block-diagonal and the i-th block is the matrix of $\mu^{(i)}$.)

Since μ commutes with the ρ_π, the restrictions commute too and $\mu^{(i)}$ is an S_n-linear map $W_i \to W_i$ (with respect to $\rho^{(i)}$). So, crucially, we can apply Schur's lemma, which tells us that $\mu^{(i)} = c_i\,\mathrm{id}_{W_i}$. Therefore, $\dim \mu_i(W_i)$ is either d_i, for $c_i \neq 0$, or 0, for $c_i = 0$.

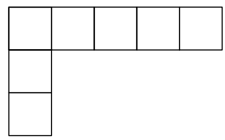
We now re-express this using characters. We have $\operatorname{Tr} \mu^{(i)} = d_i c_i$, and also $\operatorname{Tr} \mu^{(i)} = \sum_{\pi\in X_n} \operatorname{Tr} \rho^{(i)} = \sum_{\pi\in X_n} \chi^{(i)}(\pi)$, where $\chi^{(i)}$ is the character of $\rho^{(i)}$. Since all unicyclic permutations are conjugate, all the summands in the last sum are the same. So we have

$$\operatorname{rank} M = \sum_{i:\chi^{(i)}(\omega)\neq 0} d_i, \tag{31}$$

where ω is (any) unicyclic permutation.

As we know from the previous section, irreducible representations of S_n correspond to partitions of n. We need the following nontrivial fact:

Fact 5.4. *Let* χ^λ *denote the character of the irreducible representation* S^λ *(Specht module) corresponding to a partition* λ*. Then* $\chi^\lambda(\omega) \neq 0$ *iff* λ *is a "hook partition" of the form* $(\lambda_1, 1, 1, \ldots, 1)$*, such as in the next picture.*



This fact is explicitly stated, e.g., in Boerner [**Boe70**, p. 208], and it can be regarded as a special case of the *Murnaghan–Nakayama rule*, a nice algorithm for evaluating the characters of the symmetric group (also see, e.g., Sagan [**Sag01**], or Loehr [**Loe11**, Chap. 11]).

We now know that the irreducible representations corresponding to hook partitions contribute to the sum (31). Since by Corollary 3.12, the decomposition of the regular representation contains every irreducible representation with multiplicity equal to its dimension, (31) becomes

$$\operatorname{rank} M = \sum_{\lambda \text{ a hook partition}} (\dim S^\lambda)^2.$$

There are n hook partitions of n, of the form $(n-j, \underbrace{1, 1, \ldots, 1}_{j})$; here $j \in \{0, 1, \ldots, n-1\}$ is the number of squares in the "leg" of the Ferrer diagram.

As we mentioned in the previous section, $\dim S^\lambda$ is the number of standard tableaux (with increasing rows and columns) obtainable from the Ferrer diagram of λ. We must always put 1 to the upper left corner, and then, for a hook partition, we have the freedom to select j of the remaining $n-1$ numbers to go into the leg—this already determines the standard tableau. Hence $\dim S^\lambda = \binom{n-1}{j}$ and

$$\operatorname{rank} M = \sum_{j=0}^{n-1} \binom{n-1}{j}^2 = \binom{2n-2}{n-1}.$$

The latter equality can be checked via generating functions, or by double-counting the number of $(n-1)$-element subsets of $\{1, 2, \ldots, 2n-2\}$: color 1 through $n-1$ red and n through $2n-2$ blue,

and choose j red points and $n-1-j$ blue points. This concludes the proof of Theorem 5.1.

6. More Applications and Literature

Representation theory is a large and important field of modern mathematics and textbooks abound. Our exposition mostly follows Diaconis [**Dia88**, Chap. 2], which in turn is based on Serre [**Ser77**]. We have already mentioned Etingof et al. [**EGH$^+$11**] as a compact modern treatment.

Numerous impressive applications of group representations address problems in probability and statistics, often of combinatorial nature. Diaconis' book [**Dia88**] covers many of them; for example, fairly accurate estimates of how many times does one need to shuffle a deck of cards in order to get a sufficiently random ordering, for various ways of shuffling.

Shuffling is a problem concerning permutations, and representations of the symmetric group S_n are of particular significance in such cases. Here is an important and simply formulated question: given a random permutation π, let us consider the length X of the *longest increasing subsequence* of π (i.e., the largest k such that there are $i_1 < i_2 < \cdots < i_k$ with $\pi(i_1) < \pi(i_2) < \cdots < \pi(i_n)$). Elementary methods suffice to show that the expectation of X is of order $\sqrt{n}$, and with more work one can get that it is asymptotic to $2\sqrt{n}$, but what is the standard deviation, for example? A tour de force of Baik, Deift, and Johansson [**BDJ99**] determined the asymptotic distribution of X precisely; the answer is fascinating, and a part of their work relies on representations of S_n (see Aldous and Diaconis [**AD99**] for a nice exposition).

Ellis, Friedgut, and Pilpel [**EFP11**] apply representations of the symmetric group to obtain an Erdős–Ko–Rado style result about the structure of k-intersecting families of permutations (i.e., subsets of S_n in which every two permutations coincide on at least k numbers); also see Godsil and Meagher [**GM09**] for a more special result of this kind.

An amazing work of Kuperberg, Lovett, and Peled [**KLP13**] brings a new probabilistic method for establishing the existence of

interesting "rigid" combinatorial objects, such as block designs. In a part dealing with special families of permutations they also make use of representations of S_n.

Locally decodable codes are one of the miracles of computer science: When a message is encoded using such a code, there is a randomized algorithm that, given i, reconstructs the i-th bit of the original message with high probability even if a significant part, say 10%, of the encoded message is corrupted (changed). Moreover, the algorithm examines only a *constant* number of bits of the encoded message; which bits are examined depends on i and on the internal random choices made by the algorithm. Efremenko [**Efr12**] found a remarkable connection between constructions of locally decodable codes and group representations.

Püschel and Moura [**PM06**] developed a theory of algebraic algorithms for processing signals, with a significant role of group representations.

Mulmuley and coworkers (see, e.g., [**Mul11**]) launched a bold project, called *geometric complexity theory*, attempting to resolve major theoretical questions in computational complexity, such as $\mathrm{P} \neq \mathrm{NP}$ or the complexity of matrix multiplication, using advanced tools of classical mathematics, most notably algebraic geometry and representation theory. Success in bringing new complexity-theoretic results has been, let us say, limited so far, but a beautiful theory has been developed around Mulmuley's ideas. For recent contributions and good overview see, e.g., Ikenmeyer's thesis [**Ike12**].

Acknowledgment. We would like to thank Markus Püschel for an explanation concerning his work on signal processing and Clifton Ealy for reading and valuable comments.

Bibliography

[AD99] D. Aldous and P. Diaconis. Longest increasing subsequences: from patience sorting to the Baik–Deift–Johansson theorem. *Bull. Amer. Math. Soc. (N.S.)*, 36(4):413–432, 1999.

[BDJ99] J. Baik, P. Deift, and K. Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.*, 12(4):1119–1178, 1999.

[Boe70] H. Boerner. *Representations of Groups. With Special Consideration for the Needs of Modern Physics (2nd ed.)*. North-Holland Publishing Co., Amsterdam-London; American Elsevier Publishing Co., Inc., New York, 1970.

[Dia88] P. Diaconis. *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988.

[EFP11] D. Ellis, E. Friedgut, and H. Pilpel. Intersecting families of permutations. *J. Am. Math. Soc.*, 24(3):649–682, 2011.

[Efr12] K. Efremenko. From irreducible representations to locally decodable codes. In *STOC'12—Proceedings of the 2012 ACM Symposium on Theory of Computing*, pages 327–337. ACM, New York, 2012.

[EGH+11] P. Etingof, O. Golberg, S. Hensel, T. Liu, A. Schwendner, D. Vaintrob, and E. Yudovina. *Introduction to Representation Theory*, volume 59 of *Student Mathematical Library*. American Mathematical Society, Providence, RI, 2011. Online at `www-math.mit.edu/~etingof/replect.pdf`.

[GM09] C. Godsil and K. Meagher. A new proof of the Erdős–Ko–Rado theorem for intersecting families of permutations. *European J. Combin.*, 30(2):404–414, 2009.

[Ike12] C. Ikenmeyer. Geometric complexity theory, tensor rank, and Littlewood–Richardson coefficients. Ph.D. Thesis, Univ. Paderborn, 2012.

[KLP13] G. Kuperberg, S. Lovett, and R. Peled. Probabilistic existence of regular combinatorial structures. Preprint, arXiv:1302.4295, 2013.

[Loe11] N. A. Loehr. *Bijective Combinatorics*. Discrete Mathematics and its Applications (Boca Raton). CRC Press, Boca Raton, FL, 2011.

[LS93] L. Lovász and M. Saks. Communication complexity and combinatorial lattice theory. *J. Comput. System Sci.*, 47(2):322–349, 1993.

[MS82] K. Mehlhorn and E. M. Schmidt. Las Vegas is better than determinism in VLSI and distributed computing. In *Proc. 14th Annual ACM Symposium on Theory of Computing (STOC)*, pages 330–337. ACM, 1982.

[Mul11] K. D. Mulmuley. On P vs. NP and geometric complexity theory. *J. ACM*, 58(2):Art. 5, 26, 2011.

[NW95] N. Nisan and A. Wigderson. On rank vs. communication complexity. *Combinatorica*, 15(4):557–565, 1995.

[PM06] M. Püschel and J. M. F. Moura. Algebraic signal processing theory. Preprint, arXiv:cs/0612077 (later published in several journal papers), 2006.

[RS95] R. Raz and B. Spieker. On the "log rank"-conjecture in communication complexity. *Combinatorica*, 15(4):567–588, 1995.

[Sag01] B. E. Sagan. *The Symmetric Group*, volume 203 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 2001. Representations, combinatorial algorithms, and symmetric functions.

[Ser77] J.-P. Serre. *Linear Representations of Finite Groups*. Springer-Verlag, New York, 1977. Translated from the second French edition by Leonard L. Scott, Graduate Texts in Mathematics, Vol. 42.

Chapter 5

Polynomials

Here we discuss polynomials in several variables. They belong among the most powerful and most often applied mathematical tools in computer science, and sometimes their use works like a magic wand.

The set of all solutions of a system of m polynomial equations in n variables is called an *algebraic variety*, and it is studied in *algebraic geometry*, one of the most classical and deepest areas of mathematics. Here we will make the first few steps in this fascinating field.

1. Rings, Fields, and Polynomials

A *ring* R is an algebraic structure with addition and multiplication; the readers unsure about the definition might want to check it. Here we will consider only *commutative* rings (commutativity concerns multiplication, since addition in a ring is always commutative) with *unit element* 1. Unlike in usual introductory courses of algebra, we will see a large menagerie of rings.

A *field* is a ring in which we also have division (by each nonzero element, that is). We will most often consider the fields $\mathbb{R}$, the reals, and $\mathbb{C}$, the complex numbers, sometimes also a finite field $\mathbb{F}_q$ with q elements, where q, as we recall, must be a prime power, and the rationals $\mathbb{Q}$. An arbitrary field will usually be denoted by $\Bbbk$, partially

in agreement with a typical convention in algebraic geometry where k is used.

Everyone knows univariate polynomials such as $37x^5 - 2x^4 + 12$. The set of all polynomials in a variable x with coefficients in a ring R is denoted by $R[x]$. It also forms a ring, with the usual addition and multiplication of polynomials.

We will more often consider *multivariate* polynomials, such as $13x^5y^3z - 6x^2y^4z^2 + y^2 - 2z$. We write $R[x_1, \ldots, x_n]$ for the ring of all polynomials in variables $x_1, \ldots, x_n$ with coefficients in R. A polynomial $f \in R[x_1, \ldots, x_n]$ is a finite sum of **terms** of the form $c_\alpha x^\alpha$, where $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n_{\geq 0}$ is a vector of nonnegative integers, $c_\alpha \in R$ is a **coefficient**, and x^α is a convenient shorthand for the **monomial** $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. The **degree** of such a monomial is $\alpha_1 + \cdots + \alpha_n$. The degree of f, written $\deg f$, is the maximum of the degrees of its monomials.

The degree of the zero polynomial, which has no monomials, is usually taken as $-\infty$.

Each polynomial $f \in R[x_1, \ldots, x_n]$ defines a function $R^n \to R$ in an obvious way. Usually we will use the same letter for the polynomial and for the function.

Exercise 1.1. Prove (before reading further, and carefully!) that if the function defined by a polynomial $f \in \mathbb{R}[x, y]$ is zero everywhere on $\mathbb{R}^2$, then f is the **zero polynomial**; that is, all coefficients are 0. Similarly for $\Bbbk[x_1, \ldots, x_n]$ where $\Bbbk$ is an *infinite* field. (Note that if $\mathbb{F}$ is a finite field, then $\prod_{a \in \mathbb{F}}(x - a)$ is a nonzero polynomial defining the zero function $\mathbb{F} \to \mathbb{F}$.) ✠

Rigidity of polynomials. Polynomials constitute one of the most significant classes of functions, and they have various amazing properties. For computer science, one of the most important properties is some kind of *rigidity*, which can be vaguely expressed as "if two polynomials differ, then they differ *a lot*."

Here is a well-known manifestation of rigidity for univariate polynomials.

(Univariate rigidity) A nonzero univariate polynomial $f \in \Bbbk[x]$ of degree $d \geq 0$, where $\Bbbk$ is a *field*, has at most d roots. Consequently, if $f, g \in \Bbbk[x]$, $\deg(f), \deg(g) \leq d$, and $f(a) = g(a)$ for at least $d+1$ distinct points a, then $f = g$.

We recall that this is proved by induction on d, by checking that if $f(a) = 0$, then f is divisible by $x - a$.

2. The Schwartz–Zippel Theorem

This is a manifestation of rigidity in the multivariate case, one which is quite simple to prove and extremely useful, e.g., for randomized algorithms.

Theorem 2.1 (The Schwartz–Zippel theorem)**.** *Let $\Bbbk$ be a field, let d be a natural number, and let S be a finite subset of $\Bbbk$. Then for every nonzero polynomial $f \in \Bbbk[x_1, \ldots, x_n]$ of degree d, the number of n-tuples $(r_1, r_2, \ldots, r_n) \in S^n$ with $f(r_1, \ldots, r_n) = 0$ is at most $d|S|^{n-1}$. In other words, if $r_1, \ldots, r_n \in S$ are chosen independently and uniformly at random, then the probability of $f(r_1, \ldots, r_n) = 0$ is at most $\frac{d}{|S|}$.*

Here we measure the size of the zero set of f discretely, by counting the points of its intersection with the "combinatorial cube" S^n. If $\Bbbk = \mathbb{F}_q$, we can often simply take $S = \mathbb{F}_q$.

Proof of the Schwartz–Zippel theorem. We proceed by induction on n. The $n = 1$ case was mentioned earlier, so let $n > 1$.

Let us suppose that x_1 occurs in at least one term of f with a nonzero coefficient (if not, we rename the variables). Let us write f as a polynomial in x_1 with coefficients being polynomials in $x_2, \ldots, x_n$:

$$f = \sum_{i=0}^{k} f_i x_1^i, \quad f_1, \ldots, f_k \in \Bbbk[x_2, \ldots, x_n],$$

where k is the maximum exponent of x_1 in f.

We divide the n-tuples $(r_1, \ldots, r_n)$ with $f(r_1, \ldots, r_n) = 0$ into two classes. The first class, called V_1, consists of the n-tuples with

$f_k(r_2,\ldots,r_n)=0$. Since the polynomial $f_k(x_2,\ldots,x_n)$ is not identically zero and has degree at most $d-k$, the number of choices for $(r_2,\ldots,r_n)$ is at most $(d-k)|S|^{n-2}$ by the induction hypothesis, and so $|V_1|\le(d-k)|S|^{n-1}$.

The second class V_2 are the remaining n-tuples, that is, those with $f(r_1,\ldots,r_n)=0$ but $f_k(r_2,\ldots,r_n)\neq 0$. Here we count as follows: r_2 through r_n can be chosen in at most $|S|^{n-1}$ ways, and if $r_2,\ldots,r_n$ are fixed with $f_k(r_2,\ldots,r_n)\neq 0$, then r_1 must be a root of the univariate polynomial $g(x_1)=f(x_1,r_2,\ldots,r_n)$. This polynomial has degree (exactly) k, and hence at most k roots. Thus $|V_2|\le k|S|^{n-1}$, which gives $d|S|^{n-1}$ altogether, finishing the proof. □

Exercise 2.2. Check that the Schwartz–Zippel theorem is tight; i.e., find an n-variate polynomial of degree d whose zero set in S^n has $d|S|^{n-1}$ points (where $d<|S|$). ✠

A well known "continuous" counterpart of the Schwartz–Zippel theorem asserts that for every nonzero polynomial $f\in\mathbb{R}[x_1,\ldots,x_n]$, the zero set is Lebesgue null.[1] This follows, e.g., from *Sard's theorem* of mathematical analysis, or it can be proved directly.

Exercise 2.3. Imitate the proof of the Schwartz–Zippel theorem to show that the zero set of f is Lebesgue null for every nonzero $f\in\mathbb{R}[x_1,\ldots,x_n]$. (Fubini's theorem helps with a convenient proof, if you are somewhat familiar with measure theory.) ✠

3. Polynomial Identity Testing

Testing perfect matchings. We recall that a **matching** in a graph G is a set of edges $F\subseteq E(G)$ such that no vertex of G is incident to more than one edge of F. A **perfect matching** is a matching covering all vertices. One of the most famous uses of the Schwartz–Zippel theorem is in an algebraic algorithm for testing the existence of a perfect matching in a given graph.

For simplicity, we will discuss only the bipartite case. So we consider a bipartite graph, with vertices divided into two classes

[1]We recall that a set $X\subseteq\mathbb{R}^n$ is *Lebesgue null* if, for every $\varepsilon>0$, X can be covered by at most countably many axis-parallel boxes of total volume at most ε. Also see Chapter 1.

$\{u_1, u_2, \dots, u_n\}$ and $\{v_1, v_2, \dots, v_n\}$ and the edges going only between the two classes. Both of the classes have the same size, for otherwise, there is no perfect matching. Let m stand for the number of edges of G.

Let S_n be the set of all permutations of the set $\{1, 2, \dots, n\}$. Every perfect matching F of G uniquely corresponds to a permutation $\pi \in S_n$; we can write $F = \{\{u_1, v_{\pi(1)}\}, \{u_2, v_{\pi(2)}\}, \dots, \{u_n, v_{\pi(n)}\}\}$.

We express the existence of a perfect matching by a determinant, but not of an ordinary matrix of numbers, but rather of a matrix whose entries are *variables*. We introduce a variable x_{ij} for every edge $\{u_i, v_j\} \in E(G)$ (so we have m variables altogether), and we define an $n \times n$ matrix A by

$$a_{ij} := \begin{cases} x_{ij} & \text{if } \{u_i, v_j\} \in E(G), \\ 0 & \text{otherwise.} \end{cases}$$

The determinant of A is a polynomial in the m variables x_{ij}. By the definition of the determinant,

$$\begin{aligned} \det A &= \sum_{\pi \in S_n} \operatorname{sgn}(\pi) \cdot a_{1,\pi(1)} a_{2,\pi(2)} \cdots a_{n,\pi(n)} \\ &= \sum_{\substack{\pi \text{ describes a perfect} \\ \text{matching of } G}} \operatorname{sgn}(\pi) \cdot x_{1,\pi(1)} x_{2,\pi(2)} \cdots x_{n,\pi(n)}. \end{aligned}$$

Clearly, if G has no perfect matching, then $\det A$ is the zero polynomial. But the converse also holds: if G does have a perfect matching, then $\det A \neq 0$ as a polynomial. To see this, we fix a permutation σ that defines a perfect matching, and we set $x_{i,\sigma(i)} := 1$ for every $i = 1, 2, \dots, n$, while the remaining x_{ij} are set to 0. Then all terms in the above expansion of $\det A$ vanish *except* for the one corresponding to σ, which is ± 1.

So testing for a perfect matching in G is equivalent to testing if $\det A$ is the zero polynomial. We cannot afford to compute $\det A$ explicitly in the usual form, as a sum of monomials, since it may have up to $n!$ terms.

But if we substitute specific numbers for the variables x_{ij}, we can calculate the determinant reasonably fast, e.g., by Gaussian elimination. So we can imagine that $\det A$ is available through a black box, from which we can obtain its value at any specified point.

Since $\deg(\det A) \leq n$, the Schwartz–Zippel theorem shows that if $\det A$ is nonzero and we compute it for values of the x_{ij} chosen independently at random from $S := \{1, 2, \ldots, 2n\}$, then the probability of getting 0 is at most $\frac{1}{2}$. This gives a probabilistic algorithm for testing the existence of a perfect matching in G.

The probability of error can be reduced, either by repeating the algorithm several times, or by choosing from a larger set S.

Computationally, instead of working over the integers, it is better to compute the determinant over a sufficiently large finite field, because then we need not worry about the intermediate values in the computation getting very large. (There *is* a polynomial-time version of the Gaussian elimination over the integers, but it is not an easy matter.)

If we compute the determinant by Gaussian elimination, the running time is $O(n^3)$, which is worse than for some combinatorial algorithms for perfect matchings. But using fast matrix multiplication, the determinant can be computed faster; the current best asymptotic running time[2] is $O(n^{2.373})$. This yields the asymptotically fastest known perfect matching algorithm.

This algorithm also has a fast implementation on a parallel computer, with polylogarithmic running time. No other known approach yields comparably fast parallel algorithms.

Finally, it is worth mentioning that although the basic version of the algorithm, as described above, only decides if there is a perfect matching but does not find one, there are more sophisticated extensions that also find a perfect matching, and if a perfect matching does not exist, they can find a matching of maximum cardinality. See [**Har09**] for recent results and references.

---

[2] A straightforward way of multiplying two $n \times n$ matrices takes time of order n^3. The main conceptual message is that this is not optimal, as was discovered by Strassen; his algorithm is taught in basic computer-science sources. The state of art at the time of writing is Le Gall [**Gal14**] with the bound given in the text.

Counting compositions. The strategy in the above algebraic algorithm is very general and can be used for an arbitrary *polynomial identity testing*; that is, for a polynomial of controlled degree provided by a black box, the Schwartz–Zippel theorem allows us to test whether the polynomial is identically zero. Here is another lovely application.

Given a set $P \subseteq S_n$ of permutations, we want to count $|P \circ P|$, i.e., the number of distinct permutations ρ that can be expressed as a composition $\sigma\tau$ for $\sigma, \tau \in P$. Mainly for notational simplicity, let us assume $|P| = n$.

A straightforward algorithm for computing $|P \circ P|$ takes every pair $(\sigma, \tau) \in P^2$, computes the composition $\sigma\tau$ in $O(n)$ time, and then counts the number of distinct permutations in the resulting list. With some care, this can be implemented in a total of $O(n^3)$ time.

To get an asymptotically faster, algebraic algorithm, we introduce variables $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$. Let us observe that, given permutations σ and τ, the (quadratic) polynomial

$$f_{\sigma\tau} := \sum_{i=1}^{n} x_{\sigma(i)} y_{\tau^{-1}(i)}$$

encodes the composition $\rho := \sigma\tau$, in the sense that

$$f_{\sigma\tau} = \sum_{i=1}^{n} x_{\rho(i)} y_i,$$

as is easy to check. Consequently, $f_{\sigma\tau}$ and $f_{\sigma'\tau'}$ are equal polynomials iff $\sigma\tau = \sigma'\tau'$. Hence, $|P \circ P|$ equals the number of distinct polynomials among $(f_{\sigma\tau} : \sigma, \tau \in P)$.

Next, we observe that all of the $f_{\sigma\tau}$ can be evaluated simultaneously using a matrix product. Indeed, let us enumerate $P = \{\sigma_1, \ldots, \sigma_n\}$, and define the polynomial matrices A, B with $a_{ij} = x_{\sigma_j(i)}$ and $b_{ij} = y_{\sigma_j^{-1}(i)}$. Setting $C = A^T B$, we find that $c_{ij} = f_{\sigma_i \sigma_j}$.

The probabilistic algorithm for computing $|P \circ P|$ now goes as follows. We set $N := 4n^4$, $S := \{1, 2, \ldots, N\}$, we pick values $s_1, \ldots, s_n$ and $t_1, \ldots, t_n$ from S independently and uniformly at random, we make the substitutions $x_i := s_i$ and $y_i := t_i$, $i = 1, 2, \ldots, n$, and we compute the value of C. By fast matrix multiplication this can be

done in $O(n^{2.373})$ time. We return the number of distinct entries of the resulting matrix as the answer.

Clearly, this answer is never larger than $|P \circ P|$. If it is strictly smaller than $|P \circ P|$, it means that a nonzero polynomial of the form $f_{\sigma\tau} - f_{\sigma'\tau'}$ evaluates to 0 at $s_1, \ldots, s_n$, $t_1, \ldots, t_n$. For every fixed fourtuple $(\sigma, \tau, \sigma', \tau') \in P^4$, this has probability at most $\frac{1}{2n^4}$ according to the Schwartz–Zippel theorem (with degree $d = 2$). The probability that this occurs for at least one of the at most n^4 fourtuples is thus at most $\frac{1}{2}$.

Hence the answer is correct with probability at least $\frac{1}{2}$. As before, this probability can be boosted by repetition and/or by choosing larger S.

4. Interpolation, Joints, and Contagious Vanishing

We begin with a small counting question, whose result appears very often when dealing with polynomials.

Fact 4.1. *The number of monomials of degree at most d in variables $x_1, \ldots, x_n$ equals $\binom{d+n}{n}$.*

Indeed, the number in question is the number of ordered n-tuples $(\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n_{\geq 0}$ with $\alpha_1 + \cdots + \alpha_n \leq d$, and counting them is basic combinatorics which we omit here.

Somewhat imprecisely, we can say that $\binom{d+n}{n}$ is the number of "degrees of freedom" for a general n-variate polynomial of degree at most d. In a more sophisticated language, if $P_d \subset \Bbbk[x_1, \ldots, x_n]$ is the vector space (over $\Bbbk$) consisting of all polynomials of degree at most d, then all monomials of degree at most d form a basis, and so $\dim P_d = \binom{d+n}{n}$ by Fact 4.1.

The next simple, but surprisingly useful, lemma can be regarded as a kind of counterpart of the Schwartz–Zippel theorem: that theorem says that the zero set of a polynomial cannot be too big, and the lemma tells us that, nevertheless, we can cover a significant number of points by such a zero set.

Lemma 4.2. *Let $a_1, a_2, \ldots, a_N$ be points in $\Bbbk^n$, where $N < \binom{d+n}{n}$. Then there exists a nonzero polynomial $f \in \Bbbk[x_1, \ldots, x_n]$ of degree at most d such that $f(a_i) = 0$ for all i.*

Proof. Given the a_i, we regard the coefficients c_α of the desired polynomial f as unknowns. So we have $\binom{d+n}{n}$ unknowns. A requirement of the form $f(a) = 0$ translates to a *homogeneous* linear equation for the c_α. Since $N < \binom{n+d}{n}$, we have fewer equations than unknowns, and such a homogeneous system always has a nonzero solution. So there is a polynomial with at least one nonzero coefficient.

Expressed differently, we can consider the linear map $P_d \to \Bbbk^N$ that sends a polynomial f to the N-tuple $(f(a_1), \ldots, f(a_N))$. Since $\dim P_d > N$, this map has a nontrivial kernel. □

Exercise 4.3. (a) We recall that real numbers $\xi_1, \ldots, \xi_n$ are **algebraically independent** (over the rationals) if there is no nonzero polynomial $f \in \mathbb{Q}[x_1, \ldots, x_n]$ with $f(\xi_1, \ldots, \xi_n) = 0$. Prove that for every n there exist n algebraically independent real numbers. Hint: one can use a cardinality argument or a measure argument, for example.

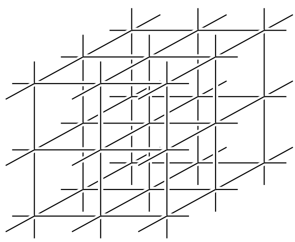
(b) Show that if $a_1, \ldots, a_N \in \mathbb{R}^n$ are points whose nN coordinates are algebraically independent, and if $N = \binom{d+n}{n}$, then the only polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$ of degree at most d vanishing at all the a_i is identically zero. ✠

Exercise 4.4. (a) Given $a_1, \ldots, a_N \in \Bbbk^n$ and values $b_1, \ldots, b_N \in \Bbbk$, prove that there exists a polynomial $f \in \Bbbk[x_1, \ldots, x_n]$ with $f(a_i) = b_i$ for all $i = 1, \ldots, N$, and with $\deg f \le N - 1$.

(b) Show that the bound $\deg f \le N - 1$ is optimal in the worst case (i.e., find $a_1, \ldots, a_N$ and $b_1, \ldots, b_N$ for which no f of smaller degree will do). Note that for $n \ge 2$, this bound is very different from the one in Lemma 4.2. ✠

The joints problem. We consider a set L of n lines in $\mathbb{R}^3$, and call a point $a \in \mathbb{R}^3$ a *joint* if there are at least three lines of L, not all coplanar, passing through a. The question is, what is the maximum possible number of joints for n lines?

There is a lower bound of $\Omega(n^{3/2})$ attained by a grid of lines,



and it was proved by Guth and Katz [**GK10**] in 2008, after many years of effort by a number of people and many intermediate results, that, asymptotically, this is the most one can get.

Theorem 4.5. *The maximum number of joints of n lines in $\mathbb{R}^3$ is $O(n^{3/2})$.*

There is a straightforward generalization to $\mathbb{R}^d$: for every fixed d, the maximum number of joints of n lines in $\mathbb{R}^d$ is of order $n^{d/(d-1)}$, where a joint means a point common to at least d lines whose direction vectors span $\mathbb{R}^d$. For simplicity we stick to the $d = 3$ case.

On partial derivatives. Let $f \in \mathbb{R}[x_1, \ldots, x_n]$ be a real polynomial. We recall that the **partial derivative** $\partial f/\partial x_i$ is the usual derivative of a univariate real function, where x_i is regarded as a variable, while all the other x_j are considered constant. The **gradient** of f is the n-tuple

$$\nabla f := \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right).$$

As a side remark, we note that the derivative can be defined purely formally, by setting $\partial(x^i)/\partial x := ix^{i-1}$ and extending linearly, and this makes sense over any field. Many of the usual properties of derivatives can then be checked as well, and so one need not specialize to real (or complex) numbers. For finite fields, though, it may be better to work with the *Hasse derivative*, where the m-th Hasse derivative $D^m(x^i) := \binom{i}{m} x^{i-m}$; this avoids troubles with dividing by zero, e.g., in a Taylor expansion formula.

If $f \in \mathbb{R}[x_1, \ldots, x_n]$ is of degree $d \geq 1$, then each of the partial derivatives $\partial f/\partial x_i$ is a polynomial of degree at most $d-1$, and at least one of them is nonzero.

The following observation connects the definition of a joint to algebra.

Observation 4.6. *Let a be a joint of lines ℓ_1, ℓ_2, ℓ_3 in $\mathbb{R}^3$, and let $f \in \mathbb{R}[x_1, x_2, x_3]$ be a polynomial that vanishes on each of the ℓ_i. Then $\nabla f(a) = 0$; that is, all of the partial derivatives of f vanish at a.*

Proof. This follows easily using the notion and simple properties of directional derivatives. Here is a more explicit argument.

W.l.o.g. we may assume $a = 0$ (the general case follows by translation). If we write $f = c_0 + c_1x_1 + c_2x_2 + c_3x_3 + (\textit{terms of degree} \geq 2)$, then we have $\frac{\partial f}{\partial x_i}(0) = c_i$, and $\nabla f(0) = c := (c_1, c_2, c_3)$. Letting $v_i = (v_{i1}, v_{i2}, v_{i3})$ be the directional vector of ℓ_i, the restriction of f to the line ℓ_i can be regarded as the univariate polynomial

$$f(tv_i) = c_0 + (c_1v_{i1} + c_2v_{i2} + c_3v_{i3})t + (\textit{terms of degree} \geq 2).$$

Thus, vanishing of this polynomial means, in particular, that $c_1v_{i1} + c_2v_{i2} + c_3v_{i3} = 0$; that is, the vector c is perpendicular to v_i. A vector perpendicular to three linearly independent vectors in $\mathbb{R}^3$ must be zero. □

We are almost ready for the proof of the $O(n^{3/2})$ upper bound for joints, but there is still a simple technical step to be prepared. If we have a set L of n lines with a large number m of joints, then an average line contains "many" joints, namely, $3m/n$. But for the polynomial argument to work, we want that *every* line contains many joints. This is taken care by a standard "pruning" argument (if you know the proof of the statement that every graph of average degree 2δ contains a subgraph with minimum degree at least δ, then you also know the proof of the next lemma).

Lemma 4.7. *Let L be a set of n lines in $\mathbb{R}^3$, let J be the set of all joints of L, and let $b := m/2n$, where $m = |J|$. Then there are subsets $L' \subseteq L$ and $J' \subseteq J$ such that $L' \neq \emptyset$, every point of J' is a joint of the lines of L', and every line of L' contains more than b points of J'.*

Proof. We use the following *pruning* procedure: We set $J_0 = J$, $L_0 = L$, and for $i = 1, 2, 3, \ldots$, if L_{i-1} contains a line ℓ with at most b joints of J_{i-1}, we set $L_i := L_{i-1} \setminus \{\ell\}$, and $J_i := J_{i-1} \setminus \ell$ (i.e., all joints in which ℓ participated are removed).

By definition, this procedure finishes with some $L' = L_k$ and $J' = J_k$ such that each point of J' is a joint of the lines in L' and each line of L' contains more than b joints of J'.

It remains to verify that $L' \neq \emptyset$, for which it suffices to check $J' \neq \emptyset$. Since we have removed at most n lines and at most b joints per line, we have $|J'| \geq m - nb = m/2 > 0$. □

Now we can focus on the essence.

Proof of Theorem 4.5. For contradiction, we suppose that a set L of n lines has $m \geq 7n^{3/2}$ joints. Let J, $b = m/2n$, J', and L' be as in the previous lemma.

We choose a nonzero polynomial $f \in \mathbb{R}[x_1, x_2, x_3]$ that vanishes on all of J' and, subject to this condition, has the *smallest possible degree.*

First we claim that $\deg f \leq b$. Indeed, by Lemma 4.2, $\deg f$ does not exceed the smallest integer d with $\binom{d+3}{3} > |J'|$, and a simple calculation shows that $\binom{b+3}{3} > m \geq |J'|$. Namely,

$$\binom{b+3}{3} > \frac{b^3}{3!} = \frac{(m/2n)^3}{3!} = m\frac{m^2}{48n^3} \geq m,$$

since we assumed $m \geq 7n^{3/2}$.

The restriction of f on each line $\ell \in L'$ is thus a univariate polynomial of degree at most $\deg f \leq b$ that vanishes in at least $b + 1$ points, and hence f vanishes everywhere on ℓ. By Observation 4.6, we have all the partial derivatives $\partial f / \partial x_i$ zero on all of J'. At the same time, since $\deg f \geq 1$, at least one of these partial derivatives is a nonzero polynomial.

But then such a nonzero partial derivative is a polynomial of degree strictly smaller than f vanishing on J', and this contradicts our choice of f and concludes the proof. □

This kind of argument is what Larry Guth calls *contagious vanishing*: the vanishing of f spreads like infection from J' to the lines of L'. In more complicated proofs of this kind, this spreading may continue further, to suitable planes or surfaces, and sometimes ultimately to the whole space.

There are several other beautiful applications of the contagious vanishing argument. The most significant ones are probably a near-optimal solution to the *Erdős distinct distances problem* due to Elekes, Guth, and Katz, and a proof of the *Kakeya conjecture over finite fields* due to Dvir. These, and much more, can be found, e.g., in Guth's book [**Gut13**] in preparation or in Tao's survey [**Tao13**]. There are also older arguments in number theory, due to Thue (see [**Gut13**]) and, especially, due to Baker (see [**Wal79**, Sec. 4]), which use some sort of contagious vanishing.

5. Varieties, Ideals, and the Hilbert Basis Theorem

Varieties. Let $\mathcal{F} \subseteq \Bbbk[x_1, \ldots, x_n]$ be a set of polynomials, possibly infinite. The **variety** $V(\mathcal{F})$ of $\mathcal{F}$ is the set of common zeros of the polynomials in $\mathcal{F}$:

$$V(\mathcal{F}) := \{(a_1, \ldots, a_n) \in \Bbbk^n : f(a_1, \ldots, a_n) = 0 \text{ for all } f \in \mathcal{F}\}.$$

Some sources use Z instead of V, Z for "zero set."

An **(algebraic) variety**[3] is any subset of $\Bbbk^n$ that can be expressed as $V(\mathcal{F})$ for some $\mathcal{F}$. More precisely, such a set is called an **affine** algebraic variety, to distinguish it from a *projective* algebraic variety, to be mentioned later.

Exercise 5.1. Show that a finite union, as well as an arbitrary intersection, of varieties is a variety. ✠

Exercise 5.2. Prove that the sets $\mathbb{Z} \subseteq \mathbb{R}$ and $[0,1]^2 \subseteq \mathbb{R}^2$ are not algebraic varieties (over $\mathbb{R}$). ✠

[3] In some of the sources an algebraic variety is also required to be irreducible (this is a notion defined later), while an arbitrary $V(\mathcal{F})$ is called an *algebraic set*.

Algebraic geometry. The study of algebraic varieties is called *algebraic geometry.* In the literature, one can encounter (at least) two quite distinct branches of algebraic geometry, with different flavor and conventions.

Classical algebraic geometry mainly investigates varieties over the complex numbers and, more generally, over **algebraically closed** fields (these are fields in which every nonconstant polynomial has a root). It is an enormous, very important, highly developed, and sometimes very abstract area of modern mathematics. Actually, since the work of Grothendieck in the 1960s, "true" algebraic geometers no longer consider algebraic varieties, but rather *schemes.* A scheme is a more general and technically convenient notion, for which even the definition is out of our scope; see, e.g., [**Gat13**] for an introduction.

Real algebraic geometry considers varieties over $\mathbb{R}$ and, more generally, **semialgebraic sets**, which are defined not only by conjunctions of polynomial equations, but also by Boolean combinations of polynomial inequalities. One can say that the results are perhaps less elegant than those about varieties over algebraically closed fields, but sometimes they are closer to the needs of computer science and other applications.

We will see a sample of basic results in both of these directions.

Ideals. We recall that an **ideal** in a (commutative) ring R is a subset $I \subseteq R$ that contains 0 and is closed under addition and under multiplication by arbitrary elements of R (in symbols, $f, g \in I$ implies $f + g \in I$ and $f \in I$, $h \in R$ implies $hf \in I$).

Exercise 5.3. Show that a ring R (commutative and with 1) has only two ideals, $\{0\}$ and R, iff it is a field. ✠

For a subset $\mathcal{F} \subseteq R$, we let $\langle \mathcal{F} \rangle$ be the **ideal generated by $\mathcal{F}$**. By definition, this is the intersection of all ideals in R that contain $\mathcal{F}$, and it is easy to check that $\langle \mathcal{F} \rangle = \{h_1 f_1 + \cdots + h_n f_n : n \geq 0, f_1, \ldots, f_n \in \mathcal{F}, h_1, \ldots, h_n \in R\}$ (this is similar to linear combinations in linear algebra, but here we multiply by arbitrary elements of R).

Specializing this to the polynomial ring $\Bbbk[x_1, \ldots, x_n]$, it is easy to see that for every set $\mathcal{F}$ of polynomials we have $V(\mathcal{F}) = V(\langle \mathcal{F} \rangle)$.

Therefore, every variety X is the set of common zeros of some ideal I in $\Bbbk[x_1, \ldots, x_n]$; $X = V(I)$. Ideals are usually much better to work with than arbitrary sets of polynomials.

Here is the first significant general result about varieties: we can restrict ourselves to *finitely generated* ideals. A ring R is called **Noetherian** if every ideal in R is generated by a finite set. In particular, every field $\Bbbk$ is Noetherian, since the only ideals are $\{0\}$ and $\Bbbk = \langle 1 \rangle$.

In the literature, the definition is often stated in a different but equivalent way: R is Noetherian iff there is no infinte sequence of properly nested ideals $I_1 \subsetneq I_2 \subsetneq \cdots$ in R.

Exercise 5.4. Check this equivalence. ✠

Theorem 5.5 (Hilbert basis[4] theorem)**.** *If R is a Noetherian ring, then the polynomial ring $R[x]$ is Noetherian as well. Consequently, $\Bbbk[x_1, \ldots, x_n]$ is Noetherian for every field $\Bbbk$.*

Hilbert's proof, more than 100 years old, was unusual at that time since it was nonconstructive: it proved the existence of a finite generating set in every ideal of $\Bbbk[x_1, \ldots, x_n]$, but did not provide any method for finding one. This nonconstructive approach was initially criticized, but later on embraced enthusiastically by the mathematical community. In the last decades, with renewed emphasis on computations and algorithms, people again put much effort into finding constructive, and efficient, proofs for important results.

Proof. Let $I \subset R[x]$ be an ideal. We are going to choose a sequence $f_1, f_2, f_3, \ldots$ of elements (polynomials) from I inductively: for $i = 1, 2, 3, \ldots$, f_i is an element of the smallest degree in $I \setminus \langle f_1, \ldots, f_{i-1} \rangle$. For $i = 1$, in particular, we have $\langle \emptyset \rangle = \{0\}$, and so f_1 is a smallest-degree nonzero element of I. We have $\deg f_1 \leq \deg f_2 \leq \cdots$.

If we reach some n with $\langle f_1, \ldots, f_n \rangle = I$, we are done.

Otherwise, let $a_i \in R$ be the **leading coefficient** of f_i (the coefficient of the highest power of x), and let us consider the ideal

[4] Here "basis" refers to what we call a "generating set." In linear algebra, bases are inclusion-minimal generating sets and they have a number of neat properties, such as all having the same size for a given vector space. In contrast, different inclusion-minimal generating sets of an ideal may have very different sizes and thus, for example, they are unsuitable for defining "dimension."

$L = \langle a_1, a_2, \ldots \rangle$. Since R is Noetherian, L is generated by $a_1, \ldots, a_m$ for some finite m.

We claim that $I' := \langle f_1, \ldots, f_m \rangle = I$. If not, then f_{m+1} was chosen as a smallest-degree element in $I \setminus I'$. The leading coefficient a_{m+1} of f_{m+1} belongs to L and thus it can be written as $a_{m+1} = \sum_{i=1}^{m} h_i a_i$, where $h_1, \ldots, h_m \in R$.

Using this, we can construct a polynomial $g \in I'$ that has the same degree and same leading coefficient as f_{m+1}, namely,

$$g := \sum_{i=1}^{m} h_i f_i x^{\deg f_{m+1} - \deg f_i}$$

Then $f_{m+1} - g$ has degree strictly smaller than f_{m+1} and lies in $I \setminus I'$ (why?). But this contradicts our choice of f_{m+1} as a smallest-degree element. □

Exercise 5.6. For every n, find an ideal in $\mathbb{R}[x, y]$ that needs at least n generators. ✠

6. The Nullstellensatz

The German word *Nullstellensatz*, meaning "zero locus theorem," is commonly used in English to denote a basic and classical result of algebraic geometry. It applies to varieties over an algebraically closed field, most notably over $\mathbb{C}$—a very important assumption.

For a field $\Bbbk$ that is not algebraically closed, one can sometimes obtain useful information by applying the Nullstellensatz with the **algebraic closure** $\overline{\Bbbk}$ of $\Bbbk$, which is an inclusion-minimal algebraically closed field extending $\Bbbk$; as it turns out, $\overline{\Bbbk}$ is determined uniquely up to isomorphism.

Exercise 6.1. (a) Prove that for every field $\Bbbk$, possibly finite, there are infinitely many irreducible polynomials in $\Bbbk[x]$, none a constant multiple of another. (Recall that a polynomial f is **irreducible** if it is not a product $f = gh$ with $\deg g, \deg h \geq 1$.)

(b) Deduce that every algebraically closed field is infinite. ✠

The weak Nullstellensatz: a theorem about alternative. There are several ways of stating the Nullstellensatz. The following

one is perhaps the most intuitive. It is called "weak" but the full version can be derived from it fairly quickly.

Many areas of mathematics have *theorems of alternative*, with the following structure: *if something cannot be done, then this impossibility must be caused by an "obvious" obstacle.* In linear algebra, for example, if a system $Ax = b$ of linear equations has no solution, then there is a linear combination of the equations that has all coefficients on the left-hand side zero and the right-hand side nonzero. In other words, there exists a vector y such that $y^T A = 0$ and $y^T b \neq 0$.

Exercise 6.2. Prove this using suitable theorems from linear algebra. ✠

The weak Nullstellensatz can also be stated in this form: if a system of polynomial equations $f_1 = f_2 = \cdots = f_m = 0$, with $f_1, \ldots, f_m \in \Bbbk[x_1, \ldots, x_n]$ and $\Bbbk$ *algebraically closed*, has no solution, then there are polynomials $h_1, \ldots, h_m \in \Bbbk[x_1, \ldots, x_n]$ such that $h_1 f_1 + \cdots + h_m f_m = 1$. The last equation is an obvious reason of unsolvability of the original system, since any common zero of $f_1, \ldots, f_m$ would also be a zero of $h_1 f_1 + \cdots + h_m f_m$, but the latter is never zero.

Here is the usual, formally somewhat different, statement.

Theorem 6.3 (Weak Nullstellensatz)**.** *Let $\Bbbk$ be algebraically closed and let I be an ideal in $\Bbbk[x_1, \ldots, x_n]$ such that $V(I) = \emptyset$; that is, there is no common zero. Then $I = \langle 1 \rangle = \Bbbk[x_1, \ldots, x_n]$.*

Exercise 6.4. (a) Give an example of how this fails over $\mathbb{R}$.

(b) Prove the weak Nullstellensatz for $n = 1$. ✠

The usual proofs of the weak Nullstellensatz, including those given here, are nonconstructive—they do not provide the h_i for given f_i. Algorithmic methods exist as well, and we will mention them later on. But it should be said that although the weak Nullstellensatz provides an "obvious" reason, or proof, of unsolvability of a given polynomial system, that proof is not necessarily very compact. Indeed, examples are known in which the smallest possible degree of the h_i has to be exponential in n (see [**Kol88**] for precise bounds).

The ideal–variety correspondence: the (strong) Nullstellensatz. The strong Nullstellensatz basically says that, over an algebraically closed field, algebraic varieties in $\Bbbk^n$ are in one-to-one correspondence with ideals in $\Bbbk[x_1, \dots, x_n]$. Or actually, not with all ideals but **radical** ones, where an ideal I is radical if $f^s \in I$ for some natural number s implies $f \in I$.

This extra condition is needed since, e.g., the ideals $\langle x \rangle$ and $\langle x^2 \rangle$ in $\mathbb{C}[x]$ both define the same variety, namely $\{0\}$—but only the first one is radical. For an arbitrary ideal I in a ring R, its **radical** $\sqrt{I}$ is defined in the expected way, as $\{f \in R : f^s \in I \text{ for some } s\}$.

Exercise 6.5. Check that $\sqrt{I}$ is an ideal. ✠

For a set $S \subseteq \Bbbk^n$, let

$$I(S) := \{f \in \Bbbk[x_1, \dots, x_n] : f \text{ vanishes on } S\};$$

clearly, this is an ideal.

Exercise 6.6. (a) Check that $V(I(X)) = X$ for every variety X, over any field. (b) Verify that $\sqrt{I} \subseteq I(V(I))$ for every ideal $I \subseteq \Bbbk[x_1, \dots, x_n]$, again over any field. ✠

Theorem 6.7 (Strong Nullstellensatz)**.** *Let $\Bbbk$ be an algebraically closed ring and let I be an ideal in $\Bbbk[x_1, \dots, x_n]$. Then $I(V(I)) = \sqrt{I}$. Thus, if $f_1, \dots, f_m \in \Bbbk[x_1, \dots, x_n]$ are polynomials and g is a polynomial that vanishes on $V(f_1, \dots, f_m)$, then then there are an integer s and polynomials $h_1, \dots, h_m \in \Bbbk[x_1, \dots, x_n]$ such that $g^s = \sum_{i=1}^m h_i f_i$.*

Proof of the strong Nullstellensatz from the weak one. The main idea is known as the *Rabinowitsch trick*: we add a new variable and a new equation to get an unsatisfiable system, for which we apply the weak Nullstellensatz.

Namely, let $I = \langle f_1, \dots, f_m \rangle$; then the polynomials $f_1, \dots, f_m$ and $x_{n+1}g - 1 \in \Bbbk[x_1, \dots, x_{n+1}]$ have no common zero in $\Bbbk^{n+1}$. So by the weak Nullstellensatz we have

$$h_1 f_1 + \cdots + h_m f_m + h_{m+1}(x_{n+1}g - 1) = 1 \tag{32}$$

for some $h_1, \dots, h_{m+1} \in \Bbbk[x_1, \dots, x_{n+1}]$.

This equality holds for every value of the variables, and in particular, with $x_{n+1} = 1/g(x_1, \ldots, x_n)$ whenever $g(x_1, \ldots, x_n) \neq 0$. Hence the rational function resulting by substituting $x_{n+1} = 1/g(x_1, \ldots, x_n)$ into the left-hand side of (32) equals 1 whenever $g \neq 0$.

We multiply both sides of the resulting equality by g^s, where s is the highest power of x_{n+1} appearing in (32). This yields the following equality of polynomial functions $\mathbb{k}^n \to \mathbb{k}$

$$h'_1 f_1 + \cdots + h'_m f_m = g^s \tag{33}$$

which holds at all points except possibly at the zeros of g (here $h'_1, \ldots, h'_m \in \mathbb{k}[x_1, \ldots, x_n]$; also note that the term $h_{m+1}(x_{n+1}g - 1)$ vanishes). Using the fact that every algebraically closed field is infinite (Exercise 6.1) and, for example, the Schwartz–Zippel theorem, we get that (33) holds as an equality of polynomials, and this concludes the proof. □

The strong Nullstellensatz shows that, with $\mathbb{k}$ algebraically closed, an algebraic variety in $\mathbb{k}^n$ and a radical ideal in $\mathbb{k}[x_1, \ldots, x_n]$ are just two ways of looking at the same object. Such alternative views of mathematical objects are often very useful.

Several proofs of the Nullstellensatz are known, usually with numerous variations. Here we essentially follow a particularly simple proof from [**Arr06**], in which we meet a classical tool—resultants.

6.1. Intermezzo: Resultants. Resultants are useful for several purposes. They provide a way of detecting when two polynomials have a nonconstant common factor (or, over an algebraically closed field, a common root), and they are useful for eliminating variables from a polynomial system or, in geometric terms, for projecting an algebraic variety on a coordinate subspace. Here we introduce them briefly, aiming mainly at the properties we will really use.

Excluding common zero of two polynomials. For a while we will be dealing with *univariate* polynomials f and g; first let us assume that they are over a field $\mathbb{k}$. Let us write them as $f(x) = \sum_{i=0}^{k} f_i x^i$ and $g(x) = \sum_{j=0}^{\ell} g_j x^j$.

To see how one can naturally arrive at the resultant, let us consider the system of two polynomial equations $f = 0$, $g = 0$. A possible way of showing that it is unsolvable, i.e., f and g have no common root, is to find polynomials $a, b \in \Bbbk[x]$ such that the polynomial $af+bg$ is a nonzero constant, say 1.

First we observe that if such a and b exist, we may as well assume $\deg a < \ell$ and $\deg b < k$. This is because if some $af + bg = 1$, then also $a'f+b'g = 1$, where $a' = a+pg$ and $b' = b-pf$ for some $p \in \Bbbk[x]$. Hence we can reduce a modulo g to have degree smaller than ℓ, and then b must have degree smaller than k, for otherwise, we would have $\deg bg \geq k + \ell > \deg af$, and so $af + bg = 1$ would be impossible.

Let us regard the coefficients of a and b as above as unknowns. The requirement $af + bg = 1$ is an equality of polynomials of degree at most $k + \ell$. By comparing the coefficients of each of the relevant powers of x on both sides, we obtain a system of $k+\ell$ linear equations with $k+\ell$ unknowns. The reader may want to write this system down and see that its matrix looks as follows (we show it for the special case $k = 5$ and $\ell = 3$, which makes clear what the general case is):

$$\begin{pmatrix} f_0 & 0 & 0 & g_0 & 0 & 0 & 0 & 0 \\ f_1 & f_0 & 0 & g_1 & g_0 & 0 & 0 & 0 \\ f_2 & f_1 & f_0 & g_2 & g_1 & g_0 & 0 & 0 \\ f_3 & f_2 & f_1 & g_3 & g_2 & g_1 & g_0 & 0 \\ f_4 & f_3 & f_2 & 0 & g_3 & g_2 & g_1 & g_0 \\ f_5 & f_4 & f_3 & 0 & 0 & g_3 & g_2 & g_1 \\ 0 & f_5 & f_4 & 0 & 0 & 0 & g_3 & g_2 \\ 0 & 0 & f_5 & 0 & 0 & 0 & 0 & g_3 \end{pmatrix}.$$

This is called the **Sylvester matrix** of f and g.

The **resultant** of f and g, denoted by $\operatorname{Res}(f, g, x)$, is the determinant of the Sylvester matrix, which is an element of $\Bbbk$. From the above discussion it is clear that if $\operatorname{Res}(f, g, x) \neq 0$, then the considered linear system has a solution, and so the desired a and b with $af + bg = 1$ exist, witnessing the nonexistence of a common root of f and g.

Exercise 6.8. (a) Using Euclid's algorithm, check that if $f, g \in \Bbbk[x]$ have no nonconstant common factor, then there are polynomials $u, v \in \Bbbk[x]$ with $uf + vg = 1$. (The reverse implication is obvious.)

(b) Using (a), prove that for $f, g \in \Bbbk[x]$, where $\Bbbk$ need not be algebraically closed, $\mathrm{Res}(f, g, x) = 0$ implies that f and g have a nonconstant common factor. ✠

Resultant over a ring. We will need a slightly more general setting, where $f, g \in R[x]$ are polynomials over a ring R (commutative with 1 as usual). The definition above still makes sense and $\mathrm{Res}(f, g, x)$ is an element of R. The next lemma, which we will need later, provides another way of showing that if $\mathrm{Res}(f, g, x) \neq 0$, then f and g have no common root.

Lemma 6.9. *For every $f, g \in R[x]$, $\deg f = k$, $\deg g = \ell$, there are $a, b \in R[x]$ with $\deg a \leq \ell - 1$, $\deg b \leq k - 1$, and $\mathrm{Res}(f, g, x) = af + bg$.*

Proof. We do the following row operations on the Sylvester matrix: for $i = 2, 3, \ldots, k + \ell$ we add the i-th row multiplied by x^{i-1} to the first row. After that the first row is

$$(f, xf, \ldots, x^{\ell-1}f, g, xg, \ldots, x^{k-1}g).$$

Expanding this determinant, which still equals $\mathrm{Res}(f, g, x)$, according to the first row, we obtain precisely an expression of the desired form $af + bg$ with a, b as in the lemma. □

6.2. Proof of the Weak Nullstellensatz. We need a lemma saying that in a polynomial of degree d, we can make the coefficient of x_1^d nonzero by a suitable invertible linear substitution. This result is quite simple; it is a special case of a more intricate result known as the *Noether normalization lemma.*

Lemma 6.10. *If $\Bbbk$ is an infinite field and $f \in \Bbbk[x_1, \ldots, x_n]$ is a polynomial of degree $d \geq 1$, then there are $\lambda_1, \ldots, \lambda_{n-1} \in \Bbbk$ such that the coefficient of x_n^d in the polynomial $f'(x_1, \ldots, x_n) := f(x_1 + \lambda_1 x_n, \ldots, x_{n-1} + \lambda_{n-1} x_n, x_n)$ is nonzero.*

Proof. Let f_d denote the sum of all terms of degree d in f (this is called the *homogeneous component* of f of degree d). Then the

coefficient of x_n in f' equals $f_d(\lambda_1, \ldots, \lambda_{n-1}, 1)$. Since $\Bbbk$ is infinite, the nonzero polynomial $f_d(x_1, \ldots, x_{n-1}, 1)$ cannot vanish everywhere on $\Bbbk^{n-1}$. □

Proof of the weak Nullstellensatz. We establish the contraposition. We assume that I is an ideal properly contained in $\Bbbk[x_1, \ldots, x_n]$, and we want to find a common zero $(a_1, \ldots, a_n) \in \Bbbk^n$ of all polynomials in I.

We proceed by induction on n, considering the case $n = 1$ settled (Exercise 6.4). So let $n > 1$.

By Lemma 6.10 we can make a change of variables so that I contains a polynomial g of degree $d \geq 1$ with the term x_n^d. Since this substitution is invertible, if we find a common zero for the ideal obtained from I after the substitution, we can convert it back to a common zero for the original I. So we assume we have $g \in I$ as above.

Let I' be the set of all polynomials in I that do not contain the variable x_n (that is, there is no term with nonzero coefficient and nonzero power of x_n). We can regard I' as a subset of $\Bbbk[x_1, \ldots, x_{n-1}]$; then it is a proper ideal (right?), and so by the inductive hypothesis, there is $(a_1, \ldots, a_{n-1})$, a common zero of all polynomials in I'.

Now we claim that the set

$$J := \{f(a_1, \ldots, a_{n-1}, x_n) : f \in I\},$$

which is obviously an ideal, is not all of $\Bbbk[x_n]$. Once we prove this claim, we will be done, since by the 1-dimensional weak Nullstellensatz all polynomials in J have a common zero $a \in \Bbbk$, and then $(a_1, \ldots, a_{n-1}, a)$ is a common zero for I.

To prove the claim, we need to check that $1 \notin J$, so for contradiction, we assume that there is $f \in I$ such that $f(a_1, \ldots, a_{n-1}, x) = 1$ (this is an equality of univariate polynomials). We fix f, as well as g as above, i.e., of degree d and with term x_n^d.

Let us consider f and g as polynomials in x_n: $f = \sum_{i=0}^{k} f_i x_n^i$, $g = \sum_{j=0}^{d} g_j x_n^j$, $f_0, \ldots, f_k, g_0, \ldots, g_d \in R := \Bbbk[x_1, \ldots, x_{n-1}]$.

By Lemma 6.9, the resultant $\mathrm{Res}(f, g, x_n) \in R$ can be written as $af + bg$ with $a, b \in R[x]$, and hence it belongs to I'. To finish the

proof, we will show that $\operatorname{Res}(f, g, x_n)$ is nonzero at $(a_1, \ldots, a_{n-1})$, and hence it cannot belong to I'.

The equality $f(a_1, \ldots, a_{n-1}, x) = 1$ means that $f_0(a_1, \ldots, a_{n-1}) = 1$ and f_1 through f_k vanish at $(a_1, \ldots, a_{n-1})$. Also, by the choice of g, we have $g_d = 1$ (identically). Looking at the Sylvester matrix of f and g, again for notational simplicity in the particular case $\deg f = 5$, $\deg g = 3$, i.e.,

$$\begin{pmatrix} f_0 & 0 & 0 & g_0 & 0 & 0 & 0 & 0 \\ f_1 & f_0 & 0 & g_1 & g_0 & 0 & 0 & 0 \\ f_2 & f_1 & f_0 & g_2 & g_1 & g_0 & 0 & 0 \\ f_3 & f_2 & f_1 & g_3 & g_2 & g_1 & g_0 & 0 \\ f_4 & f_3 & f_2 & 0 & g_3 & g_2 & g_1 & g_0 \\ f_5 & f_4 & f_3 & 0 & 0 & g_3 & g_2 & g_1 \\ 0 & f_5 & f_4 & 0 & 0 & 0 & g_3 & g_2 \\ 0 & 0 & f_5 & 0 & 0 & 0 & 0 & g_3 \end{pmatrix},$$

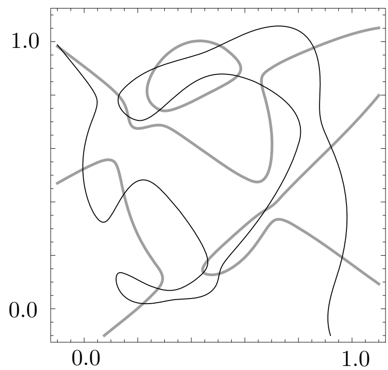
we see that at $(a_1, \ldots, a_{n-1})$ it is an upper triangular matrix with 1s on the main diagonal, and hence $\operatorname{Res}(f, g, x_n)(a_1, \ldots, a_{n-1}) = 1$. □

7. Bézout's Inequality in the Plane

One of the questions that often comes up in applications is, given a system of polynomial equations $f_1 = 0, \ldots, f_m = 0$, $f_1, \ldots, f_m \in \Bbbk[x_1, \ldots, x_n]$, what can be said about the existence and number of solutions? In order to avoid trivialities, we always assume that $d_i := \deg f_i \geq 1$ for all i.

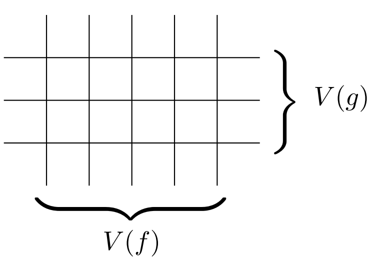
In general this is not an easy question, and in this section we will consider the special case with two equations $f(x, y) = 0$, $g(x, y) = 0$ in two variables, which is considerably simpler than the general setting but still interesting. (We are leaving aside the case of a single equation $f = 0$, which has already been treated to some extent, at least implicitly.)

Here is an example of the zero set of two polynomials f and g of degree 5; each of them has been created by passing the zero set through 25 random points in $[0, 1]^2$ using Lemma 4.2.



The system $f = g = 0$ may have infinitely many solutions—this we can see already in the case of linear equations, where f may be a multiple of g or, speaking geometrically, the two lines described by the equations may coincide. For two polynomials, infinitely many solutions may occur if f and g have a nonconstant common factor, and for $\Bbbk$ algebraically closed they actually do occur in such a case.

As we will see, excluding a common factor leads to finitely many solutions. Let us consider an example with $f(x,y) = (x-1)(x-2)\cdots(x-k)$ and $g(x,y) = (y-1)(y-2)\cdots(y-\ell)$. We have $\deg f = k$, $\deg g = \ell$, and the zero sets look like this:

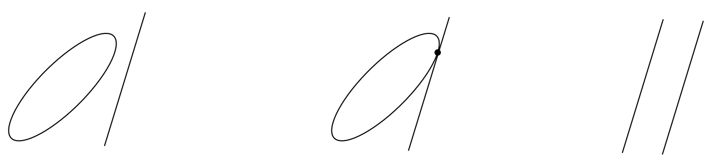


Thus, there can be as many as $k\ell$ distinct solutions. This example, trivial as it may look, is actually quite useful: the union of k hyperplanes is the zero set of a degree-k polynomial, and although this is not really a typical polynomial, it can serve for a quick sanity check of many things.

The following theorem asserts that, assuming no common factor, we cannot have more solutions than in the example.

Theorem 7.1 (Bézout's inequality in the plane). *Let $f, g \in \Bbbk[x, y]$ be polynomials of degrees $k, \ell \geq 1$, respectively, having no nonconstant common factor. Then $|V(f, g)| \leq k\ell$.*

In algebraic geometry, Bézout's theorem is often stated as an equality: under suitable assumptions, there are *exactly* $k\ell$ solutions. The assumptions have to address three issues: (a) the field has to be algebraically closed; (b) we have to count solutions with appropriately defined multiplicity; and (c) we also have to count solutions "at infinity." The next drawing illustrates these issues:



not algebraically closed   multiplicity   at infinity

We will talk about solutions at infinity later. Handling multiplicities properly takes a substantial amount of work, and we will not consider it here. However, Bézout's theorem is usually applied in the inequality form.

Theorem 7.1 can be proved in several ways, for example using resultants. The proof shown below is ingenious, short, and introduces a general approach used for handling the concept of dimension in algebraic geometry. We begin with some general considerations.

Coordinate rings, and measuring them. We recall from algebra that if I is an ideal in a (commutative) ring S, we can form the *quotient ring* S/I, whose elements are equivalence classes of elements of S, with a, b equivalent if $a - b \in I$. Here we will consider the case where S is the polynomial ring $\Bbbk[x_1, \ldots, x_n]$ and I is an ideal in S.

In particular, if $X \subseteq \Bbbk^n$ is an algebraic variety and $I = I(X) \subseteq \Bbbk[x_1, \ldots, x_n]$, then the quotient ring $\Bbbk[x_1, \ldots, x_n]/I$ is called the **coordinate ring** of X and denoted by $\Bbbk[X]$. It has an intuitive meaning: its elements can be represented by polynomials, but two polynomials are considered the same if they coincide on X (strictly speaking, this is literally true only over infinite fields).

Being determined by the ideal $I = I(X)$, the coordinate ring carries the same information as I, but some things are more convenient to express in terms of the coordinate ring. Moreover, $\Bbbk[X]$ is more suitable for representing the variety X "up to isomorphism" (which we are going to define later).

Now we want to measure the "size" of coordinate rings. Slightly more generally, we consider an ideal I and the quotient ring $R := \Bbbk[x_1, \ldots, x_n]/I$. They are both closed under addition and under multiplication by elements of $\Bbbk$, and so they are also vector spaces over $\Bbbk$.

The vector space dimension of R, or of I, in itself is usually not a very good measure of "size," since it is most often infinite. Certainly, for $I = I(X)$ and $R = \Bbbk[X]$, it does not capture the intuitive geometric notion of dimension of the variety X. The trick is to consider subspaces consisting of polynomials up to some given degree d.

For the ideal I this can be done in the obvious manner: we let $I_{\le d}$ consist of all polynomials in I of degree at most d. For R this is slightly more tricky, since two polynomials representing the same element of R may have different degrees.

We define R_d as the quotient vector space $\Bbbk[x_1, \ldots, x_n]_{\le d}/I_{\le d}$, so the elements of R_d are represented by polynomials of degree at most d, with the same equivalence as that for R.

By a well known fact from linear algebra about quotient spaces, we have $\dim R_d + \dim I_{\le d} = \dim \Bbbk[x_1, \ldots, x_n]_{\le d} = \binom{n+d}{n}$, the last equality being Fact 4.1. In particular, R_d and $I_{\le d}$ have finite dimension for every d.

The vector-space dimension of $\Bbbk[X]_d$, considered as a function of d, carries a lot of information about the variety X, and it has a name—again after Hilbert.

Let $R = \Bbbk[x_1, \ldots, x_n]/I$ be a quotient of the polynomial ring $\Bbbk[x_1, \ldots, x_n]$, and let R_d be the vector space defined as above. Then the **Hilbert function** of R (or, for $I = I(X)$, also of X) is defined as

$$\mathrm{HF}_R(d) := \dim R_d.$$

If $I \subseteq I' \subseteq \Bbbk[x_1, \ldots, x_n]$ are ideals and R, R' are the corresponding quotient rings, we have $\mathrm{HF}_R \geq \mathrm{HF}_{R'}$ (this follows from $\mathrm{HF}_R(d) = \binom{n+d}{n} - \dim I_{\leq d}$). For varieties this yields $\mathrm{HF}_X \leq \mathrm{HF}_{X'}$ for $X \subseteq X'$, which we will freely use in the sequel.

Proof of Theorem 7.1. The plan for proving the planar Bézout inequality is now this:

(i) We check that if $X \subseteq \Bbbk^2$ is an m-point set, then the Hilbert function of X is at least m for all sufficiently large d.

(ii) We show that if f and g have no nonconstant common factor, then $\mathrm{HF}_R(d) \leq k\ell$, where $R := \Bbbk[x,y]/\langle f, g\rangle$, again for sufficiently large d.

To prove (i), let $X = \{a_1, \ldots, a_m\} \subset \Bbbk^2$, and let us choose a system $\varphi_1, \ldots, \varphi_m$ of functions $X \to \Bbbk$ that are linearly independent. For example, we can set $\varphi_i(a_j) := \delta_{ij}$, the Kronecker delta, with $\delta_{ii} = 1$ and $\delta_{ij} = 0$ for $i \neq j$.

According to Exercise 4.4(a), for each φ_i there is a polynomial $p_i \in \Bbbk[x,y]$ whose values on X coincide with φ_i. Then the p_i are linearly independent as elements of the coordinate ring $\Bbbk[X]$, and this proves $\dim \Bbbk[X]_d \geq m$ for all $d \geq \max \deg p_i$. (This argument works for any number of variables, not only two.)

As for (ii), let us first consider the ideals $K := \langle f\rangle$ and $L := \langle g\rangle$. We claim that for $d \geq k$, we have $\dim K_{\leq d} = \dim \Bbbk[x,y]_{\leq d-k}$. This is because every polynomial in $p \in K$ has the form $p = af$, and p determines a uniquely. (Here we use that $\Bbbk[x,y]$ is a *unique factorization domain*; Exercise 7.3 below.) Of course, we also have $\dim L_{\leq d} = \dim \Bbbk[x,y]_{d-\ell}$ for $d \geq \ell$.

What we want to bound is $\dim I_{\leq d}$, where $I = \langle f, g\rangle$. We have $I = \{af + bg : a, b \in \Bbbk[x,y]\} = \{p + q : p \in K, q \in L\}$. The sum of two polynomials of degree at most d again has degree at most d, and hence $I_{\leq d} \supseteq K_{\leq d} + L_{\leq d}$.

Exercise 7.2. Find an example where this inclusion is proper. ✠

Fortunately, since we need to bound $\dim R_d$ from above and thus $\dim I_{\leq d}$ from below, the inclusion goes in the right direction. By the

well-known formula for the dimension of a sum of vector spaces, we have

$$\dim(K_{\le d} + L_{\le d}) = \dim K_{\le d} + \dim L_{\le d} - \dim(K \cap L)_{\le d}.$$

It remains to note that, since f and g have no common factor, a polynomial divisible by both f and g must be divisible by fg, and so $K \cap L = \langle fg \rangle$. Hence $\dim(K \cap L)_{\le d} = \dim \Bbbk[x,y]_{\le d-k-\ell}$ for $d \ge k+\ell$.

The rest is calculation with binomial coefficients:

$$\begin{aligned}
\dim R_d &= \dim \Bbbk[x,y]_{\le d} - \dim I_{\le d} \\
&\le \dim \Bbbk[x,y]_{\le d} - \dim(K_{\le d} + L_{\le d}) \\
&= \tbinom{d+2}{2} - \tbinom{d-k+2}{2} - \tbinom{d-\ell+2}{2} + \tbinom{d-k-\ell+2}{2} \\
&= k\ell
\end{aligned}$$

(assuming $d \ge k + \ell$). □

Exercise 7.3. We recall that a (commutative) ring R is called an **integral domain** if the product of every two nonzero elements is nonzero. An element $a \in R$ is *irreducible* if it cannot be written as a product $a = bc$ with neither b nor c invertible.

(a) Let R be an integral domain in which every nonzero element has a unique factorization into irreducibles (unique up to reordering and multiplication by invertible elements). The *contents* $\mathrm{cont}(f)$ of a polynomial $f \in R[x]$ is defined as the greatest common divisor of all coefficients of f. Show that $\mathrm{cont}(fg) = \mathrm{cont}(f)\,\mathrm{cont}(g)$.

(b) Prove that every univariate polynomial $f \in \Bbbk[x]$ over a field has a unique factorization into irreducible polynomials.

(c) Prove by induction on n that every $f \in \Bbbk[x_1, \ldots, x_n]$ has a unique factorization into irreducible polynomials. ✠

8. More Properties of Varieties

In this section we introduce further basic notions and results concerning algebraic varieties. Building this theory properly with all details requires much more space, and so we try to present a reasonable selection. We will encounter many clever and sophisticated notions, and one should not expect to master all of them quickly, but hopefully they will look less frightening next time. Reading this section should

give some first impression and basic vocabulary; for serious work one should study a proper textbook.

8.1. Irreducible Components. The union of the x-axis and y-axis in the plane is an algebraic variety, namely, $V(xy)$, which can naturally be decomposed into two proper subvarieties, $V(x)$ and $V(y)$. Varieties that cannot be further decomposed are called irreducible:

A variety $X \subseteq \Bbbk^n$ is **irreducible** if we cannot express $X = X_1 \cup X_2$ with X_1 and X_2 both varieties and proper subsets of X.

As we have remarked, some sources even reserve the term variety only for irreducible varieties, and irreducibility is extremely important. We have already seen a hint of this in Bézout's inequality, and many other theorems require irreducibility assumptions. For example, it turns out that an irreducible variety over an algebraically closed field has the same "local dimension" in the neighborhood of each point (we have not yet defined dimension rigorously, but surely the reader has some intuitive idea), while a reducible variety may be, e.g., the union of a plane and a line.

Exercise 8.1. (a) (Any field) Show that if a variety $X \subseteq \Bbbk^n$ is irreducible, then $I = I(X)$ is a **prime ideal**; that is, $fg \in I$ implies $f \in I$ or $g \in I$.

(b) (Any field) Prove that if $X \subseteq \Bbbk^n$ is a variety with $I(X)$ prime, then X is irreducible.

(c) Check that a prime ideal is radical, but not necessarily the other way around. ✠

Proposition 8.2. *Every variety X can be decomposed as a finite union $X = X_1 \cup \cdots \cup X_k$ of irreducible varieties. Moreover, assuming that $X_i \not\subseteq X_j$ for all $i \neq j$, the decomposition is unique up to reordering.*

The X_i as in the proposition are called the **irreducible components** of X.

Sketch of proof. Finiteness follows from the Hilbert basis theorem: if we could keep decomposing indefinitely, we would obtain an infinite descending chain of varieties $X_1 \supsetneq X_2 \supsetneq X_3 \supsetneq \cdots$, whose corresponding ideals would form an infinite ascending chain, and this is impossible since $\Bbbk[x_1, \ldots, x_n]$ is Noetherian.

As for uniqueness, assuming two minimal decompositions into irreducibles $X = X_1 \cup \cdots \cup X_k = X'_1 \cup \cdots \cup X'_\ell$, we observe that if some X_i were not among the X'_j, then the $X_i \cap X'_j$ would properly decompose X_i or vice versa. □

One of the basic sources of difficulties in algebraic geometry is that *the intersection of irreducible varieties need not be irreducible.* A simple example is with two irreducible algebraic curves in $\Bbbk^2$ intersecting in at least two points, but there are more interesting higher-dimensional examples as well, one is mentioned in Section 9 below.

We also stress that the task of finding the irreducible decomposition of a given variety is highly nontrivial in general, although algorithmically solvable.

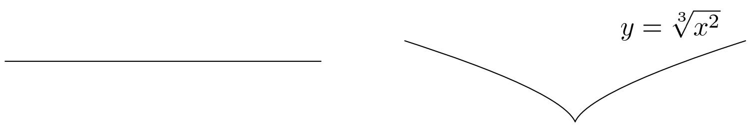
The Zariski topology. In the language of algebraic geometry, a set $S \subseteq \Bbbk^n$ is called **Zariski closed** or just **closed** if it is a variety, and it is **(Zariski) open** if its complement is a variety. Readers familiar with the notion of topological space can check that this defines a topology on $\Bbbk^n$, although a somewhat peculiar one. Nonempty open sets are very big (assuming an infinite field), they are dense in $\Bbbk^n$ and every two intersect. Thus, the topology is not Hausdorff. Yet it provides a convenient framework and terminology.

Exercise 8.3. Let $X \subseteq \Bbbk^m$ and $Y \subseteq \Bbbk^n$ be irreducible varieties. Prove that the product $X \times Y \subseteq \Bbbk^{m+n}$ is irreducible as well. ✠

8.2. Morphisms of Affine Varieties. Having defined a class of objects, affine algebraic varieties in our case, one should ask what is an appropriate notion of *morphisms* of the objects. Familiar examples of morphisms include linear maps of vector spaces, homomorphisms of groups, rings, fields, but also of graphs, and continuous maps of topological spaces.

For affine algebraic varieties, morphisms are called regular maps. A *polynomial map* $f\colon \Bbbk^m \to \Bbbk^n$ is a map $f = (f_1, \ldots, f_n)$ such that each f_i is given by a polynomial in $\Bbbk[x_1, \ldots, x_m]$. If $X \subseteq \Bbbk^m$ and $Y \subseteq \Bbbk^n$ are varieties, then a *regular map* $f\colon X \to Y$ is a map that is a restriction of a polynomial map $\bar{f}\colon \Bbbk^m \to \Bbbk^n$ to X and satisfies $f(X) \subseteq Y$.

An **isomorphism** of affine varieties is a regular map with a regular inverse. While the affine line $\mathbb{R}$ is homeomorphic as a topological space to the "cusp curve" $V(x^2 - y^3)$, it can be shown that they are not isomorphic as affine varieties.



We note that if $f\colon X \to Y$ is a regular map and $\varphi\colon Y \to \Bbbk$ is a polynomial function on Y, i.e., an element of the coordinate ring $\Bbbk[Y]$, then the composition $\varphi f\colon X \to \Bbbk$ belongs to $\Bbbk[X]$. Thus, the composition with f induces a mapping $f^*\colon \Bbbk[Y] \to \Bbbk[X]$ (note the change of direction compared to f!). Moreover, f^* is a $\Bbbk$-algebra homomorphism, meaning that it is a ring homomorphism for which, in addition, $f^*(\alpha) = \alpha$ for every $\alpha \in \Bbbk$.

Conversely, it is not hard to show that every $\Bbbk$-algebra homomorphism $\Bbbk[Y] \to \Bbbk[X]$ equals f^* for some regular map $f\colon X \to Y$.

Exercise 8.4. Prove that; start with $X = \Bbbk^m$, $Y = \Bbbk^n$. ✠

It follows that two varieties are isomorphic exactly if their coordinate rings are isomorphic as $\Bbbk$-algebras. So the coordinate ring provides a "coordinate-free" representation of a variety, independent of a specific embedding of the variety in some $\Bbbk^n$.

A useful way of proving irreducibility. Let $X \subseteq \Bbbk^m$ be an irreducible variety, and let $f\colon X \to \Bbbk^n$ be a regular map. Then it is easy to check that the image $f(X)$ is irreducible, but the statement has to be understood in a right way.

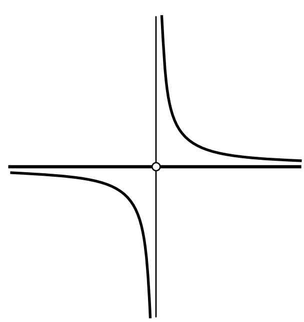
Indeed, as we will discuss below in more detail, $f(X)$ need not be a variety! So we generalize irreducibility to an arbitrary set $S \subseteq \Bbbk^n$,

meaning that we cannot write $S = (S \cap X_1) \cup (S \cap X_2)$, where X_1, X_2 are varieties and $S \cap X_1 \neq S \neq S \cap X_2$. Then we can see that if $f(X)$ were reducible, then so would be X, because the preimage of a variety under a regular map is always a variety, as is easy to check.

Thus, in particular, if we can express some variety Y parametrically, as the image of some $\mathbb{k}^m$, or of some other irreducible variety X, under a polynomial map f, then Y is irreducible. More generally, it suffices that the image $f(X)$ be **Zariski dense** in Y, meaning that Y is the smallest variety containing $f(X)$.

As an example, let m, n and $r \leq \min(m, n)$ be natural numbers, and consider the **determinantal variety** $D_r(m, n)$ consisting of all $m \times n$ matrices, considered as points in $\mathbb{k}^{mn}$, that have rank strictly smaller than r. This is indeed a variety since the rank condition can be expressed as vanishing of all $r \times r$ minors. Since an $m \times n$ matrix A has rank at most $r - 1$ iff it can be expressed as a product UV, where U is $m \times (r-1)$ and V is $(r-1) \times n$, we have a surjective regular map $\mathbb{k}^{(r-1)(m+n)} \to D_r(m, n)$, and hence the determinantal variety is irreducible.

Projections and images of affine varieties: constructible sets. Let us consider the variety $X := V(xy - 1)$, a hyperbola, and project it onto the x-axis:



The projection $\pi(X)$ is the x-axis minus 0, certainly not an algebraic variety. Passing to an algebraically closed setting, complex numbers, does not help—the 0 is still missing. So affine algebraic varieties are not closed under projections, and under regular maps in general.

One remedy is to add points at infinity and work in the projective space—see Section 8.5 below. Another approach is to consider a larger class consisting of all sets obtainable from varieties by finitely many set-theoretical operations; these are called **constructible sets**. Using the fact that varieties are closed under intersections and finite unions, it is not difficult to check that every constructible set can be written as

$$(X_1 \setminus Y_1) \cup \cdots \cup (X_k \setminus Y_k),$$

for varieties $X_1, Y_1, \ldots, X_k, Y_k$, where we may assume the X_i irreducible and $Y_i \subsetneq X_i$. Then Y_i can be regarded as a set of "exceptional points" in X_i; as we will discuss in Section 8.3, it has smaller dimension than X_i.

We state the following result without proof:

Theorem 8.5 (Chevalley's theorem)**.** *Let* $\Bbbk$ *be an algebraically closed field, and let* $\pi\colon \Bbbk^{m+n} \to \Bbbk^n$ *denote the projection on the last* n *coordinates. Then* $\pi(Z)$ *is a constructible set for every constructible set* $Z \subseteq \Bbbk^{m+n}$ *and, in particular, for every variety* Z*.*

This is actually a result about *quantifier elimination* in the theory of algebraically closed fields, and a nice proof can be found in [**MO02**].

Corollary 8.6. *The image of a constructible set* $Z \subseteq \Bbbk^m$ *under a regular map* $f\colon \Bbbk^m \to \Bbbk^n$ *is a constructible set.*

Sketch of proof. This is a generally useful trick: one needs to check that the graph $G := \{(x, f(x)) \in \Bbbk^m \times \Bbbk^n : x \in Z\}$ is a constructible set; then $f(Z) = \pi(G)$ is constructible by Chevalley's theorem. □

Exercise 8.7. The goal of this exercise is to show that, in the algebraically closed case, a "typical" projection of an affine variety is an affine variety.

(a) Let $X = V(I)$ be an affine variety in $\Bbbk^n$ defined by an ideal I, with $\Bbbk$ algebraically closed. Suppose that $f \in I$ is a polynomial of degree $d \geq 1$ with a nonzero coefficient of x_n^d. Let $\pi\colon \Bbbk^n \to \Bbbk^{n-1}$ be the standard projection given by $\pi(a_1, \ldots, a_n) = (a_1, \ldots, a_{n-1})$. Recycle some of the arguments in the proof of the weak Nullstellensatz in Section 6.2 to show that $\pi(X) = V(I')$, where $I' \subseteq \Bbbk[x_1, \ldots, x_{n-1}]$

is the ideal consisting of all polynomials in I in which the variable x_n does not occur.

Together with Lemma 6.10, this shows that for every affine variety $X \subseteq \Bbbk^n$, a generic projection of the form $(x_1, \ldots, x_n) \mapsto (x_1 + \lambda_1 x_n, \ldots, x_{n-1} + \lambda_{n-1} x_n)$ maps X onto a variety in $\Bbbk^{n-1}$.

(b) Find an example of a variety in $\mathbb{R}^2$ for which no projection, of the form given above, is a variety. ✠

Rational maps. A **rational map** $\varphi\colon \Bbbk^m \to \Bbbk^n$ is given by an n-tuple of rational functions

$$\varphi = \Big(\frac{f_1}{g_1}, \ldots, \frac{f_n}{g_n}\Big),$$

where $f_1, g_1, \ldots, f_n, g_n \in \Bbbk[x_1, \ldots, x_m]$ are polynomials, none of the g_i identically zero.

There is a catch: a rational map is not really a map in the usual sense, because it is undefined on the zero sets of the g_i (for this reason, instead of the usual mapping arrow $\to$, one uses $\dashrightarrow$ for a rational map). Nevertheless, it is defined on a Zariski open subset of $\Bbbk^m$, and it is still useful.

A rational map $\varphi\colon X \dashrightarrow Y$ of varieties, with $X \subseteq \Bbbk^m$ irreducible and $Y \subseteq \Bbbk^n$ is, similar to regular maps, a restriction of a rational map $\overline{\varphi}\colon \Bbbk^m \dashrightarrow \Bbbk^n$ to X such that $\varphi(X) \subseteq Y$, but with the extra condition that none of the denominators g_i (assuming f_i and g_i having no common factors) vanishes identically on X.

Two rational maps $X \dashrightarrow Y$ are considered equivalent if they agree on a nonempty Zariski open subset of X (they may be defined on different Zariski open subsets of X, though).

We have seen that an algebraic counterpart of regular maps $X \to Y$ are $\Bbbk$-algebra homomorphisms $\Bbbk[Y] \to \Bbbk[Y]$ of the coordinate rings. Similarly, rational maps $\varphi\colon X \dashrightarrow Y$ of *irreducible* varieties correspond to $\Bbbk$-algebra homomorphisms $\Bbbk(Y) \to \Bbbk(X)$, where $\Bbbk(X)$ is the quotient field of the coordinate ring $\Bbbk[X]$ (which is an integral domain for X irreducible, so a quotient field makes sense).

The corresponding notion of isomorphism is called *birational equivalence*, and it is more permissive than the isomorphism defined

by regular maps. For example, it is known, and not extremely difficult to prove, that every variety (over an algebraically closed field) is birationally equivalent to a *hypersurface*, i.e., a variety defined by a single polynomial.

8.3. Dimension and Degree. The dimension of algebraic varieties is defined algebraically, and it has several rather different-looking but equivalent definitions. Here we will mention only some of them, and we will not prove their equivalence.

In this section we will assume an algebraically closed field unless stated otherwise. Things are considerably subtler over an arbitrary field, and it is often preferable to work with schemes there, rather than varieties.

Dimension. Here is a definition which is very simple to state, but rather difficult to work with. The **dimension** of a variety X is the largest n such that there is a chain of properly increasing *irreducible* varieties $\emptyset \subsetneq X_0 \subsetneq \cdots \subsetneq X_n \subseteq X$. (In particular, the empty variety $\emptyset$ has dimension -1.)

The idea is that a proper subvariety of an irreducible variety must be of lower dimension; note that the same definition works for finite-dimensional vector spaces. Since, in the algebraically closed case, irreducible varieties correspond to prime ideals (Exercise 8.1), the dimension is also the length of the longest chain of properly nested prime ideals in $I(X)$ (this notion is called the *Krull dimension* of the coordinate ring of X).

With this definition, even $\dim \Bbbk^n = n$ is not obvious (but it is true).

A geometric view, and degree. Another, more geometric, way is to define the dimension of a variety $X \subseteq \Bbbk^n$ as the largest dimension k of a linear subspace $H \subset \Bbbk^n$ such that there is a projection $\pi\colon \Bbbk^n \to H$ with $\pi(X)$ Zariski dense in H. Here a *projection* is a linear map $\pi\colon \Bbbk^n \to \Bbbk^n$ such that $\pi\pi = \pi$, and $H = \pi(\Bbbk^n)$.

Another, but equivalent, geometric definition of the dimension considers only the the usual projections on all k-dimensional coordinate subspaces.

It turns out that the property of $\pi(X)$ being Zariski dense in $H = \pi(\Bbbk^n)$ is generic, in the sense that the set of the π not having this property is negligible: if we parameterize all projections π onto k-dimensional subspaces by suitable coordinates, then those with $\pi(X)$ not Zariski dense in H satisfy a nontrivial polynomial equation.

This point of view also brings us to the notion of degree. For a projection π and a point $y \in H = \pi(\Bbbk^n)$, let us consider the number of preimages $|\{x \in X : \pi(x) = y\}|$. It can be shown that for π and y generic, this number is finite and depends only on X. It is called the **degree** of X and denoted by $\deg X$.

There is also a "dual" view: if X is a k-dimensional variety in $\Bbbk^n$, then a generic $(n-k-1)$-dimensional affine subspace of $\Bbbk^n$ avoids X, while a generic $(n-k)$-dimensional affine subspace intersects it in $\deg X$ points.

Dimension and regular maps. Regular maps do not increase dimension: if X and Y are varieties (over an algebraically closed field) and $f(X) = Y$, or more generally, if $f(X)$ is Zariski dense in Y for a regular map f, then $\dim Y \le \dim X$. Moreover, if we have $\dim f^{-1}(y) = m$ for all y from a Zariski dense subset of Y, then $\dim Y = \dim X - m$. Proofs can be found in many introductory textbooks.

Generalized Bézout. If X, Y are varieties (over an algebraically closed field), then $\deg(X \cap Y) \le (\deg X)(\deg Y)$, which can be seen as a generalization of Bézout's inequality (see Heintz [**Hei83**]).

The Hilbert function and the Hilbert polynomial. We recall that the Hilbert function of a variety X is defined as the Hilbert function $\mathrm{HF}_{\Bbbk[X]}$ of its coordinate ring, and the value $\mathrm{HF}_{\Bbbk[X]}(d)$ is the dimension of the vector space $\Bbbk[X]_d$, which consists of polynomials of degree at most d modulo the polynomials in $I(X)$ of degree at most d.

It turns out that for all sufficiently large d, the Hilbert function coincides with a polynomial, called the **Hilbert polynomial** of X. More precisely, for every quotient ring $R = \Bbbk[x_1, \dots, x_n]/I$ there exist d_0 and a polynomial, denoted by HP_R and obviously uniquely determined, such that $\mathrm{HP}_R(d) = \mathrm{HF}_R(d)$ for all $d \ge d_0$.

This fact, mysterious as it may look, is not difficult. A short algebraic proof can be found, e.g., in [**Sch03**, Lemma 2.3.3], and below we will provide a geometric picture explaining the polynomial behavior.

The Hilbert polynomial provides a seemingly very different definition of dimension and degree:

> The dimension k of an affine algebraic variety X is the degree of its Hilbert polynomial $\mathrm{HP}_{\Bbbk[X]}$, and the degree of X is $k!$ times the leading coefficient of the Hilbert polynomial.

Monomial orderings. For presenting the promised geometric view of the Hilbert function, we first need to define a linear ordering of the monomials in $\Bbbk[x_1, \ldots, x_n]$; this will also be indispensable later, when we briefly discuss computational aspects of ideals and varieties.

One particular ordering that works for our purposes is the **graded lexicographic ordering**: for two monomials $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ and x^β, we first compare the degrees, i.e., $\|\alpha\|_1 = \sum_{i=1}^n \alpha_i$ and $\|\beta\|_1$, and if they are equal, we compare the nonnegative integer vectors α and β lexicographically.

More generally, a **monomial ordering** is a linear ordering $\leq$ on $\mathbb{Z}^n_{\geq 0}$ (we identify monomials with their exponent vectors) that is a well-ordering,[5] and such that $\alpha < \beta$ implies $\alpha + \gamma < \beta + \gamma$ for every $\gamma \in \mathbb{Z}^n_{\geq 0}$. For the considerations in this section, we also need the monomial ordering to be **graded**, meaning that $\|\alpha\|_1 < \|\beta\|_1$ implies $\alpha < \beta$.

So we fix a graded monomial ordering $\leq$. Then every polynomial $f \in \Bbbk[x_1, \ldots, x_n]$ has a uniquely determined **leading monomial** $\mathrm{LM}(f)$, the one that is the largest according to the monomial ordering.

For an ideal I in $\Bbbk[x_1, \ldots, x_n]$, we let $\mathrm{LM}(I) := \langle \mathrm{LM}(f) : f \in I \rangle$; this is a *monomial ideal*, meaning that it is generated by monomials (but of course, being an ideal, it also contains polynomials

---

[5] That is, every nonempty subset has a minimum element.

that are not monomials). We should also warn that if I is generated by some polynomials $f_1, \ldots, f_m$, $\mathrm{LM}(I)$ may be larger than $\langle \mathrm{LM}(f_1), \ldots, \mathrm{LM}(f_m)\rangle$—the reader may want to find an example.

The next claim, which we leave as an exercise, shows that, as far as the Hilbert function is concerned, it is enough to deal with monomial ideals.

Exercise 8.8. Let us fix a graded monomial ordering, let I be an ideal in $\Bbbk[x_1, \ldots, x_n]$, let $I' := \mathrm{LM}(I)$, and let $R := \Bbbk[x_1, \ldots, x_n]/I$ and $R' := \Bbbk[x_1, \ldots, x_n]/I'$ be the corresponding quotient rings.

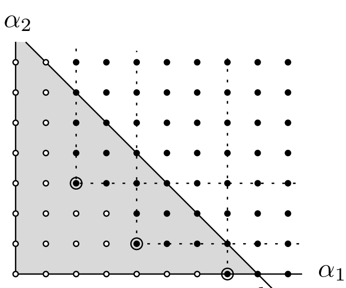
(a) Show that $I_{\le d}$ has a basis $(f_1, \ldots, f_m)$ such that $\mathrm{LM}(f_1) > \cdots > \mathrm{LM}(f_m)$, and derive $\dim I_{\le d} \le \dim I'_{\le d}$.

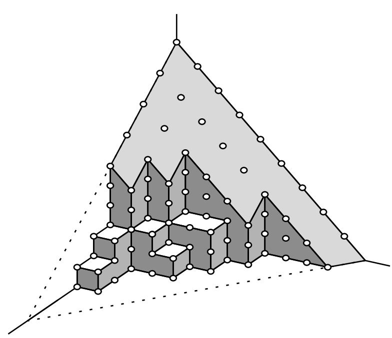
(b) Prove that if the f_i constitute a basis of $I_{\le d}$ as in (a), then $\mathrm{LM}(f_1), \ldots, \mathrm{LM}(f_m)$ generate $I'_{\le d}$. Conclude that $\mathrm{HF}_{R'} = \mathrm{HF}_R$.

(c) Where does the argument use the assumption that the monomial ordering is graded? ✠

The proof in the exercise also shows that all monomials in $I' = \mathrm{LM}(I)$ are linearly independent, and that each $I'_{\le d}$ has a basis consisting of monomials.

Let us consider $\mathbb{Z}^n_{\ge 0}$, all n-tuples of nonnegative integers, and let us color the exponent vector of every monomial in the monomial ideal I' black. This is illustrated in the left drawing for $n = 2$:





Since I' is an ideal, the black dots are the union of finitely many "corners", i.e., translations of the nonnegative orthant—one corner for each generator. The generators are marked by double circles.

The number of black dots in the halfspace $\alpha_1 + \cdots + \alpha_n \le d$ is the vector-space dimension of $I'_{\le d}$ (since the corresponding monomials form a basis), and hence the value of $\mathrm{HF}_{R'}(d)$ is the number of white dots in that halfspace (because $\mathrm{HF}_{R'}(d) = \binom{n+d}{d} - \dim I'_{\le d}$; we do *not* claim that the corresponding monomials form a basis).

From this interpretation one can see why the Hilbert function eventually becomes a polynomial: the key observation is that if we ignore a finite number of "irregular" white dots near the origin, the remaining white dots can be organized into finitely many disjoint axes-parallel "orthants" of various dimensions (semiinfinite rays, quadrants of planes, octants of 3-dimensional subspaces, etc.); this is not quite a proof but almost. The 3-dimensional picture above on the right illustrates how the halfspace $\sum \alpha_i \le d$ sweeps the set of white dots, after it has already passed the irregular part.

Finally, let us see why the growth of the Hilbert polynomial is related to the geometric dimension $V(I')$, at least for a monomial ideal I'.

Some thought reveals that $\mathrm{HP}_{R'}$ grows at least linearly iff at least one of the coordinate axes has no black dots. Assuming, e.g., that all dots on the α_1-axis are white, this means that every generator in the monomial ideal I' is a multiple of one of $x_2, \ldots, x_n$, and hence the x_1-axis is contained in $V(I')$.

Similarly, $\deg \mathrm{HP}_{R'} \ge 2$ iff there is a two-dimensional coordinate plane without a black point. Assuming it is the $\alpha_1\alpha_2$ plane, we can see that the x_1x_2-plane is contained in $V(I')$, and so on—in general, the degree of the Hilbert polynomial is the largest dimension of a coordinate subspace contained in $V(I')$. (And since I' is a monomial ideal, $V(I')$ is the union of coordinate subspaces.)

The proofs relating the Hilbert polynomial to the other definitions of dimension and degree mentioned earlier are not too difficult, but we do not treat them here.

8.4. Computation with Ideals and Gröbner bases. Here we briefly consider algorithmic questions concerning varieties and ideals.

A basic algorithmic problem is the *ideal membership*. Given an ideal $I \subseteq \Bbbk[x_1, \ldots, x_n]$, specified by a list of generators, i.e.,

$I = \langle f_1, \ldots, f_m \rangle$, how can we test whether a given polynomial g belongs to I?

Recalling that $g \in I$ means $g = \sum_{i=1}^{m} h_i f_i$ for some h_i, one way might be to look for the h_i, say by solving a system of linear equations for their coefficients. But, as we have remarked, the required degrees of the h_i may be very high, and this method is not practical.

If we have $n = 1$, i.e., univariate polynomials, every ideal can be generated by a single polynomial f, and testing whether $g \in \langle f \rangle$ is very simple: we just reduce g modulo f and see if the remainder is 0. This, of course, assumes that we know a single generator: if I is given by several generators $f_1, \ldots, f_m$, then we first need to compute their greatest common divisor.

The division algorithm. Back in the multivariate setting and trying to proceed analogously, the first question is, given generators $f_1, \ldots, f_m$, what does it mean to reduce them "modulo $f_1, \ldots, f_m$"? We would like to write $g = a_1 f_1 + \cdots + a_m f_m + r$, for suitable polynomials $a_1, \ldots, a_m$ and r, where r should be a "remainder" after the division of g by the f_i.

A good way of doing this is to fix a monomial ordering $\leq$, as introduced in the previous section (but this time it need not be graded), and always try to get rid of the leading monomial of the current g by subtracting the right multiple of some f_i.

Here is the division algorithm. It receives g as input, and successively reduces it by subtracting suitable multiples of the f_i, while simultaneously building the remainder r.

(1) Set $r := 0$.

(2) Let $\mu := \mathrm{LM}(g)$ be the leading monomial of the current g. If there is some i such that $\mathrm{LM}(f_i)$ divides μ, choose one (arbitrarily), and subtract the appropriate multiple of f_i from g so that the coefficient of μ after the subtraction is 0. Repeat this step with the new g. If there is no such i, go to the next step.

(3) At this point none of the $\mathrm{LM}(f_i)$ divides $\mathrm{LM}(g)$. Subtract the leading term of g (i.e., $\mathrm{LM}(g)$ with the coefficient it has

in g) from g and add it to r. If $g = 0$, finish, and otherwise, go back to the previous step.

This algorithm is finite, since it strictly decreases $\mathrm{LM}(g)$, according to the monomial ordering, in each step.

But, unfortunately, it is not sufficient to test ideal membership unless we have a very good set of generators. For example, if we run it with $f_1 = x^2+y$, $f_2 = xy+x$, $g = x^2-y^2$, and the graded lexicographic order as in the preceding section, we get a nonzero remainder $-y^2-y$. Yet $g \in \langle f_1, f_2\rangle$, since $x^2 - y^2 = -yf_1 + xf_2$.

The problem here is that, in the expression $-yf_1+xf_2$, the leading terms cancel out.

Gröbner bases. It turns out that, for a given monomial ordering, every polynomial ideal I has a "very good" set of generators, called Gröbner[6] basis, for which the division algorithm above is guaranteed to test membership in I correctly: it returns remainder 0 iff $g \in I$.

This can be taken as a definition of a Gröbner basis. An equivalent condition, and the usual definition, is this:

> An m-tuple $f_1, \dots, f_m$ is a **Gröbner basis** of an ideal I, with respect to a given monomial order, if $I = \langle f_1, \dots, f_m\rangle$ and $\mathrm{LM}(I) = \langle \mathrm{LM}(f_1), \dots, \mathrm{LM}(f_m)\rangle$. In words, the f_i generate I and their leading monomials generate the ideal of the leading monomials of all polynomials in I.

A Gröbner basis of I w.r.t. one monomial order may fail to be a Gröbner basis for a different monomial order.

A Gröbner basis $f_1, \dots, f_m$ is called *reduced* if it satisfies a certain natural minimality condition, namely, the leading monomials of the f_i have coefficient 1, and no monomial in any f_i is in the ideal generated by the $\mathrm{LM}(f_j)$ for $j \neq i$. For a given I and monomial order, it can be shown that a reduced Gröbner basis is unique.

There are algorithms that, given an arbitrary set of generators of I, compute a Gröbner basis, usually a reduced one, w.r.t. a given

[6] Often spelled Groebner in English texts and software.

monomial order. This algorithmic task has been investigated extensively, since it is very significant both in theory and in practice. In the worst case, the computational complexity, as well as the size of the resulting Gröbner basis, are at least exponential in n, the number of variables.

Once a Gröbner basis is available, we can solve the ideal membership problem by the division algorithm. Many other tasks can be solved as well: computing the sum, intersection, or quotient of two ideals; computing the dimension, Hilbert polynomial, and Hilbert function of a given variety; solving a system of polynomial equations; etc. The worst-case computational complexity of these problems is again very high, but the existing implementations can sometimes handle impressively large instances.

A nice mathematical application of these algorithms is for automatic theorem proving: with Gröbner bases and some cleverness one can make a computer program routinely prove many theorems in high-school geometry or even beyond it; for example, the Pappus theorem. The method is sketched in [**CLO07**].

Here we finish our very brief excursion to algorithms, referring to [**CLO07**] for a thorough introduction.

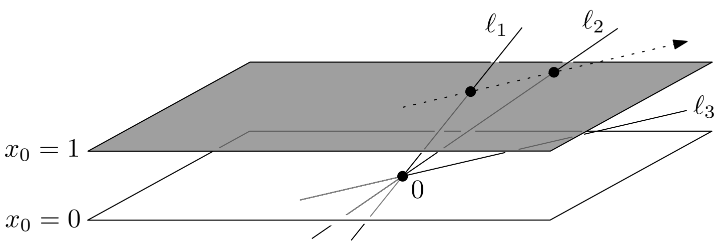
8.5. Projective Varieties. Instead of the affine space $\mathbb{k}^n$, algebraic geometry is usually done in the projective space $\mathbb{P}^n = \mathbb{P}^n(\mathbb{k})$, which can be thought of as a completion of $\mathbb{k}^n$ by adding points at infinity in a suitable way. Then almost everything comes out more elegantly and algebraic varieties behave much better—for example, over an algebraically closed field, the projection of a variety is again a variety, unlike in the affine case.

The projective space. To construct $\mathbb{P}^n$ formally, we consider all $(n+1)$-tuples $(a_0 : a_1 : \cdots : a_n)$, where $a_0, \ldots, a_n \in \mathbb{k}$ are not all simultaneously 0. Each point a of $\mathbb{P}^n$ is an equivalence class of such $(n+1)$-tuples consisting of all nonzero multiples of some $(n+1)$-tuple:

$$a = \{(\lambda a_0 : \lambda a_1 : \cdots : \lambda a_n) : \lambda \in \mathbb{k} \setminus \{0\}\}.$$

Such an equivalence class can be viewed as a line through the origin in $\Bbbk^{n+1}$. The $(n+1)$-tuple $(a_0 : \cdots : a_n)$ is called the **homogeneous coordinates** of a; these are defined only up to a scalar multiple.

The following picture illustrates, for the case $n = 2$, the geometric meaning of this construction.



Here $\Bbbk^2$, to which we want to add points at infinity, is embedded in $\Bbbk^3$ as the gray plane $x_0 = 1$ (where the coordinates in $\Bbbk^3$ are x_0, x_1, x_2 and the x_0-axis is drawn vertical). Each point a of this plane corresponds to the line $\overline{0a}$ through the origin in $\Bbbk^3$.

Conversely, each line through the origin corresponds to exactly one point of the gray plane, *except* for horizontal lines, such as ℓ_3. When we start tilting the line ℓ_1 towards the position ℓ_2 and further towards the horizontal position ℓ_3, the corresponding point in the gray plane recedes to infinity along the dashed line. So horizontal lines such as ℓ_3 correspond to points at infinity, one point for each direction of parallel lines in the gray plane.

Algebraically, in this interpretation, a point of $\mathbb{P}^n$ with homogeneous coordinates $(x_0 : \cdots : x_n)$ with $x_0 \neq 0$ corresponds to the point $(\frac{x_1}{x_0}, \ldots, \frac{x_n}{x_0}) \in \Bbbk^n$. Adding the points at infinity, with $x_0 = 0$, can be thought of as adding a copy of $\mathbb{P}^{n-1}$ to $\Bbbk^n$.

On the other hand, the structure of $\mathbb{P}^n$ is the same everywhere and locally, in the neighborhood of each point, it looks like the affine space $\Bbbk^n$. In our picture, the plane representing $\Bbbk^2$ can be rotated around 0, and this yields various ways of placing $\Bbbk^2$ in $\mathbb{P}^2$. In algebraic geometry, this allows one to transfer all kinds of "local" notions from the affine setting to the projective one.

Projective varieties. We would like to say that projective varieties are zero sets of polynomial systems of equations in $\mathbb{P}^n$, but we have to be a bit careful.

Working in $\mathbb{P}^n$, we have $n+1$ coordinates $x_0, \ldots, x_n$, but it does not make sense to consider, for example, the equation $x_1 = x_0^2$, since the $(n+1)$-tuple $(1 : 1 : \cdots : 1)$ satisfies it, but $(2 : 2 : \cdots : 2)$, representing the same point of $\mathbb{P}^n$, does not.

One has to consider only zero sets of **homogeneous polynomials** $f \in \Bbbk[x_0, \ldots, x_n]$, meaning that all monomials of f have the same degree; then the zero set can be regarded as a subset of $\mathbb{P}^n$. The counterpart for ideals is a **homogeneous ideal**, one generated by homogeneous polynomials (but necessarily containing non-homogeneous polynomials too); for such an ideal I, the variety $V(I) \subseteq \mathbb{P}^n$ is well defined as the set of common zeros of all $f \in I$.

Every polynomial $f \in \Bbbk[x_1, \ldots, x_n]$ can be *homogenized* to a homogeneous polynomial $\tilde{f}$ by adding an appropriate power of x_0 to each term so that the resulting polynomial becomes homogeneous (and has the same degree as f). For instance, from $x_1^3 + x_1x_2 + 5$ we get $x_1^3 + x_0x_1x_2 + 5x_0^3$. An ideal $I \subseteq \Bbbk[x_1, \ldots, x_n]$ is homogenized to the homogeneous ideal $\tilde{I} = \langle \tilde{f} : f \in I \rangle \subset \Bbbk[x_0, \ldots, x_n]$. (Let us mention that the homogenization of a generating set of I need not generate $\tilde{I}$.) From an affine variety $V(I) \subseteq \Bbbk^n$ we thus obtain the **projective completion** $V(\tilde{I}) \subseteq \mathbb{P}^n$ (it is perhaps worth mentioning that isomorphic affine varieties may have nonisomorphic projective completions). The meaning of $I(X)$ for $X \subseteq \mathbb{P}^n$ is also modified appropriately.

Many of the concepts and results from the affine setting transfer to projective varieties without change (irreducible decomposition, Zariski open and closed sets) or with only minor modifications.

For the weak Nullstellensatz, $V(I) = \emptyset$ not only for $I = \langle 1 \rangle$, but also when the radical of I is $\langle x_0, x_1, \ldots, x_n \rangle$. This *irrelevant ideal* also has to be excluded in the strong Nullstellensatz; after that, over an algebraically closed field, we have a bijective correspondence between homogeneous radical ideals and projective varieties.

A **morphism** $f\colon X \to Y$ of projective varieties $X \subseteq \mathbb{P}^m$ and $Y \subseteq \mathbb{P}^n$ needs to be defined locally: for each $x_0 \in X \subseteq \mathbb{P}^m$ there is a Zariski open neighborhood U and homogeneous polynomials $f_0, \dots, f_n \in \Bbbk[x_0, \dots, x_m]$ of the same degree such that $f(x) = (f_0(x) : \cdots : f_n(x))$ for all $x \in U$ (and in particular, at least one $f_i(x)$ must be nonzero for each x).

As for the Hilbert function, in the projective case one needs to take the dimension of $\Bbbk[x_0, \dots, x_n]/I_{=d}$, where $I_{=d}$ is the vector subspace spanned by homogeneous polynomials of degree *exactly* d in the homogeneous ideal I.

Cutting with a polynomial. If X is a k-dimensional projective variety over an algebraically closed field and f is a polynomial, then $k-1 \leq \dim(X \cap V(f)) \leq k$. If, moreover, X is irreducible and f does not vanish on it, then $\dim(X \cap V(f)) = k-1$.

Exercise 8.9. Show that this fails for affine varieties; $\dim(X \cap V(f))$ can be smaller than $\dim X - 1$. ✠

Projection. Unlike in the affine case, the projection of a projective variety is also a projective variety, and so is the image under a morphism.

One has to be slightly careful with what is meant by a projection, since in $\mathbb{P}^n$ we cannot simply omit some of the homogeneous coordinates, because we might get all 0s.

One way around this is to consider a projection as a map $\pi\colon \mathbb{P}^m \times \mathbb{P}^n \to \mathbb{P}^n$, but strictly speaking, we have not yet defined what a variety in $\mathbb{P}^m \times \mathbb{P}^n$ is. There are two equivalent ways of doing that.

First, we may embed $\mathbb{P}^m \times \mathbb{P}^n$ as a variety in $\mathbb{P}^{(m+1)(n+1)-1}$; this is called the **Segre embedding**, and it sends a pair $((x_0 : \cdots : x_m), (y_0 : \cdots : y_n))$ to $(x_0y_0 : x_1y_0 : \cdots : x_iy_j : \cdots : x_my_n)$. Then varieties in $\mathbb{P}^m \times \mathbb{P}^n$ are just the intersections of varieties in $\mathbb{P}^{(m+1)(n+1)-1}$ with the embedded copy of $\mathbb{P}^m \times \mathbb{P}^n$. (We note in passing that the image of the Segre embedding is essentially the determinantal variety $D_2(m+1, n+1)$ mentioned in Section 8.2.)

Second, and more explicitly, a variety in $\mathbb{P}^m \times \mathbb{P}^n$ is the common zero set of a set of *bihomogeneous* polynomials

$$f \in \Bbbk[x_0, \dots, x_m, y_0, \dots, y_n],$$

where f is bihomogeneous if each monomial has degree k in the x_i and degree ℓ in the y_i, for some k, ℓ, possibly with $k \neq \ell$. Then the result can be stated as follows:

Theorem 8.10 (Projection theorem)**.** *For every projective variety $Z \subseteq \mathbb{P}^m \times \mathbb{P}^n$ over an algebraically closed field, $\pi(Z)$ is also a projective variety, where $\pi\colon \mathbb{P}^m \times \mathbb{P}^n \to \mathbb{P}^n$ is the projection onto the second factor.*

Let us prove at least something in this long section.

Proof. Let $f_1, \dots, f_r$ be bihomogeneous generators of $I(Z)$. We may assume that all of them have the same degree k in the x_i. (If, in order to achieve that, we need to raise the degree of f_1 by d, say, we replace f_1 by the $(d+1)$-tuple of polynomials $x_0^d f_1, x_1^d f_1, \dots, x_n^d f_1$, which does not change the zero set.)

Let us fix a point $a \in \mathbb{P}^m$ and write $f_{i,a}(x) := f_i(x, a)$. By definition, $a \notin \pi(Z)$ means that the $f_{i,a}$ have no common zero in $\mathbb{P}^n$. By the projective weak Nullstellensatz, this happens iff the radical of the homogeneous ideal $I := \langle f_{1,a}, \dots, f_{r,a} \rangle \subseteq \Bbbk[x_0, \dots, x_m]$ contains the irrelevant ideal $\langle x_0, \dots, x_m \rangle$. In other words, there are $s_0, \dots, s_m$ with $x_i^{s_i} \in I$, $i = 1, 2, \dots, m$.

For the proof to work, we transform this condition further: setting $s := s_0 + \cdots + s_m$, we can see that I contains *all* homogeneous polynomials of degree s. So $a \notin \pi(Z)$ if and only if there exists s such that for every homogeneous $g \in \Bbbk[x_0, \dots, x_m]$ of degree s we can find $h_1, \dots, h_r \in \Bbbk[x_0, \dots, x_m]$ with

$$g = \sum_{i=1}^{r} h_i f_{i,a}. \tag{34}$$

Here, crucially, since all the $f_{i,a}$ are homogeneous of degree k, we may assume that the h_i are homogeneous of degree $s - k$, because monomials of any other degree can be discarded from them without changing the validity of (34).

Therefore, for every g, (34) can be rewritten as a system of linear equations for the unknown coefficients of the h_i. The matrix of this system, call it A, does not depend on g, and its entries are homogeneous polynomials in $a_0, \ldots, a_n$, the homogeneous coordinates of a. The number of equations is t, the number of monomials of degree s in $m+1$ variables; it equals $\binom{s+m}{m}$ but we do not need that.

The solvability of (34) for every g means that the linear system is solvable for every right-hand side, which means exactly that A has rank t. Hence the *negation* of this condition can be expressed as vanishing of all the $t \times t$ minors of A.

Let Y_s be the set of all $a \in \mathbb{P}^n$ such that the matrix A, defined above, has rank less than t. Each Y_s is a variety, and we have $\pi(Z) = \bigcap_{s=0}^{\infty} Y_s$. Therefore, $\pi(Z)$ is a projective variety as claimed. □

9. Bézout's Inequality in Higher Dimensions

9.1. In Search of a Proper Statement. We again consider the system of polynomial equations $f_1 = 0, \ldots, f_m = 0$, $f_1, \ldots, f_m \in \mathbb{k}[x_1, \ldots, x_n]$, this time for $n > 2$ variables.

The most important case is $m = n$. Guided by the example with hyperplanes, i.e., with $f_i = (x_i - 1) \cdots (x_i - d_i)$, where $d_i = \deg f_i \geq 1$, we expect that if the number of solutions is finite, then it should be at most $d_1 d_2 \cdots d_n$. Moreover, finitely many solutions should be the typical, "generic" case.

Warning example. Unlike in the planar case, over an arbitrary field, having finitely many solutions does *not* guarantee that the bound $d_1 d_2 \cdots d_n$ for their number is correct.

Indeed, the system of three equations

$$(x-1)^2(x-2)^2 \cdots (x-k)^2 + (y-1)^2(y-2)^2 \cdots (y-k)^2 = 0, \ z = 0, \ z = 0$$

has k^2 solutions in $\mathbb{R}^3$, but the degrees are $2k, 1, 1$. We note that the solution set in $\mathbb{C}^3$ is infinite.

Another example. In the previous example, the first equation has only a 1-dimensional solution set over $\mathbb{R}$, while the two remaining equations are identical. However, over $\mathbb{C}$ the solution set of the first

equation is 2-dimensional, and so, at least over algebraically closed fields, one might hope to exclude this kind of pathology by imposing a suitable condition on the f_i. Indeed, drawing inspiration from the planar case, a natural guess for such a condition can be that no two of the f_i have a common factor.

However, things are not that simple, and the suggested condition is definitely not the right one. Here is a highly instructive example for $n = 3$:

$$f_1 = x^3 - yz, \ \ f_2 = y^2 - xz, \ \ f_3 = z^2 - x^2y.$$

These are irreducible polynomials, as is easy to check, none a multiple of another. But $V(f_1, f_2, f_3)$ contains the curve C with parametric expression

$$C = \{(t^3, t^4, t^5) : t \in \mathbb{C}\},$$

and so surely it is not finite.

This example is also interesting in another respect. In linear algebra, every k-dimensional vector subspace of $\mathbb{k}^n$ can be described by $n - k$ linear equations; for example, a line in $\mathbb{R}^3$ is always the intersection of two planes. In contrast, the curve C *cannot* be defined by two polynomial equations: It is easy to check the common zero set of every two of the f_i contains points not belonging to the zero set of the third—e.g., $V(f_1, f_2)$ contains the z-axis, where f_3 is nonzero. With more effort, one can show that *no two* polynomials suffice; this is done algebraically, by checking that the ideal $\langle f_1, f_2, f_3 \rangle$ cannot be generated by two polynomials.

Let us remark that things cannot get completely out of hand with examples of this kind: it is known that every irreducible affine variety in $\mathbb{k}^n$, $\mathbb{k}$ algebraically closed, can be given as the zero set of at most $n + 1$ polynomials [**Hei83**, Prop. 3].

Bézout's inequality assuming finitely many zeros. It seems that there is no particularly useful general condition for $V(f_1, \ldots, f_n)$ to be finite, although there are algorithms that can decide this question for any given $f_1, \ldots, f_n$—but these are nontrivial and quite demanding computationally.

One way around this is to *assume* $V(f_1, \ldots, f_n)$ is finite. Then, for $\Bbbk$ algebraically closed, the expected inequality for the number of zeros does hold.

Theorem 9.1 (Higher-dimensional Bézout's inequality I)**.** *Let* $\Bbbk$ *be algebraically closed, and let* $f_1, \ldots, f_n \in \Bbbk[x_1, \ldots, x_n]$ *be polynomials of degrees* $d_1, \ldots, d_n \geq 1$*. Assuming that* $V(f_1, \ldots, f_n) \subset \Bbbk^n$ *is* finite*, it has at most* $d_1 d_2 \cdots d_n$ *points.*

Actually, one can say a bit more: even if $V(f_1, \ldots, f_n)$ contains irreducible components of positive dimension, the number of one-point irreducible components is still at most $d_1 d_2 \cdots d_n$.

We will not prove Theorem 9.1 here. A reasonably accessible algebraic proof can be found in [**Tao12**, Sec. 8.4].

Bounding the number of nonsingular zeros. The above formulation of Bézout's inequality leaves something to be desired, since, as we have mentioned, verifying the assumption $|V(f_1, \ldots, f_n)| < \infty$ is not easy in general (although there are various sufficient conditions known; see, e.g., [**CLO05**, Chap. 3,4] and [**Sch95**]).

Another formulation, which is often useful for applications, is to consider only a suitable kind of "nice" zeros; namely, only those where the hypersurfaces $X_i := V(f_i)$ intersect transversally.

We will work only over the field $\mathbb{R}$, where one can rely on intuition and methods from analysis. However, with an appropriate generalization of notions like gradient, results can also be obtained for other fields—see [**CKW11**, Sec. 5].

Let $X_1, \ldots, X_n \subseteq \mathbb{R}^n$ be the hypersurfaces as above and let a be a point where they all intersect. Transversality means that if we make a tangent hyperplane h_i to each X_i at a, then these n hyperplanes intersect only in a—they look like the coordinate hyperplanes, after a suitable affine transformation (this includes the assumption that each X_i is $(n-1)$-dimensional in some neighborhood of a).

We recall that if $f\colon \mathbb{R}^n \to \mathbb{R}$ is a differentiable function, then the gradient ∇f at a point a is the "fastest ascent" direction for f. Assuming $f(a) = 0$, $\nabla f(a)$ is perpendicular to the zero set of f, and thus it is a normal vector of the tangent hyperplane of the zero

set, assuming $\nabla f(a) \neq 0$. (Rigorously this can be derived from the implicit function theorem.)

The transversality of our $X_1, \ldots, X_n$ at a thus corresponds to linear independence of the n gradients $\nabla f_1(a), \ldots, \nabla f_n(a)$, or in other words, to the **Jacobian determinant**

$$J_{f_1,\ldots,f_n}(a) := \det \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \ldots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & \ldots & \vdots \\ \frac{\partial f_n}{\partial x_1}(a) & \ldots & \frac{\partial f_n}{\partial x_n}(a) \end{pmatrix}$$

being nonzero. (Apologies to the readers for whom the geometric meaning of the Jacobian is well known and boring.)

A point $a \in V(f_1, \ldots, f_n)$ with $J_{f_1,\ldots,f_n}(a) \neq 0$ is called a **nonsingular zero**.

Theorem 9.2 (Higher-dimensional Bézout's inequality II)**.** *Let $f_1, \ldots, f_n \in \mathbb{R}[x_1, \ldots, x_n]$. Then the polynomial system $f_1 = 0, \ldots, f_n = 0$ has at most $d_1 d_2 \cdots d_n$ nonsingular zeros in $\mathbb{R}^n$, where $d_i = \deg f_i$.*

9.2. Proof for Nonsingular Zeros. We present a proof of Theorem 9.2 due to Wooley [**Woo96**], mostly following a presentation in [**CKW11**, Sec. 5]. For another more or less elementary proof, going via a complex version of the theorem, see [**BPR03**, Sec. 4.7].

So let $a_1, \ldots, a_N \in \mathbb{R}^n$ be nonsingular common zeros of $f_1, \ldots, f_n$; we want to show that $N \leq D = d_1 d_2 \cdots d_n$.

First we fix a linear polynomial $\pi \in \mathbb{R}[x_1, \ldots, x_n]$ such that the $\pi(a_i)$ are all distinct; we can think of this as choosing a projection on a suitable line. Armed with the knowledge from the previous sections, the reader will surely supply a rigorous proof of existence of a suitable π.

The general idea of the proof is to produce a nonzero univariate polynomial of degree at most D for which all the $\pi(a_i)$ are roots.

To this end, we would like to have a polynomial $h \in \mathbb{R}[y_1, \ldots, y_n, z]$ satisfying the following conditions:

(C1) The polynomial $\tilde{h} := h(f_1, \ldots, f_n, \pi) \in \mathbb{R}[x_1, \ldots, x_n]$, obtained by substituting $f_i(x_1, \ldots, x_n)$ for y_i and $\pi(x_1, \ldots, x_n)$ for z into h, is the zero polynomial.

(C2) The highest power of z occurring in h is at most D.

(C3) The univariate polynomial $h_0(z) := h(0, 0, \ldots, 0, z)$ is not identically zero.

If we had such an h, we would be done: by (C1), $h_0(\pi(a)) = 0$ whenever a is a common zero of the f_i, by (C2) we have $\deg h_0 \le D$, and together with (C3) this would show that h_0 has at most D zeros.

However, a suspicious thing is that this plan does not use the nonsingularity of the considered common zeros of the f_i, and indeed, we will have to modify it. But (C1) and (C2) can be achieved; this is done by linear algebra and counting, and it works over any field.

Lemma 9.3. *Given arbitrary polynomials* $f_1, \ldots, f_n, \pi \in \Bbbk[x_1, \ldots, x_n]$ *with* $\deg f_i = d_i$ *and* $\deg \pi = 1$*, there exists a nonzero polynomial* $h \in \Bbbk[y_1, \ldots, y_n, z]$ *satisfying (C1) and (C2).*

We postpone the proof of the lemma. Having such an h, we cannot guarantee (C3), unfortunately. But here we use the assumption with nonsingular zeros to perturb the f_i, and for the perturbed version we will be able to get (C3).

Concretely, we perturb by choosing a sufficiently small vector $\delta = (\delta_1, \ldots, \delta_n) \in \mathbb{R}^n$ and considering the perturbed system $f_1 = \delta_1, \ldots, f_n = \delta_n$. We claim that if a is a nonsingular zero of the original system, with zero right-hand sides, then for every δ sufficiently small, there is a zero $a(\delta)$ of the perturbed system, such that $a(\delta) \to a$ as $\|\delta\| \to 0$.

This is a textbook application of the implicit function theorem; after all, a nonzero Jacobian is typically used this way. We just consider the function $F\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ given coordinate-wise by $F(x, \delta)_i := f_i(x) - \delta_i$. Then $F(a, 0) = 0$, and the implicit function theorem guarantees the existence of a (continuous) function $a(\delta)$ with $F(a(\delta), \delta) = 0$ for all δ sufficiently small (note that the Jacobian in the implicit function theorem is with respect to the "dependent"

variables, which in our case are the x_i, and this is exactly $J_{f_1,\ldots,f_n}(a)$ as in the definition of nonsingular zero).

It follows that if the original system has at least N nonsingular zeros, so does the perturbed system for δ sufficiently small. Moreover, again for δ small enough, these N zeros of the perturbed system still yield N distinct values of the projection π. So if h satisfies (C1) and (C2), then for every $\delta \in \mathbb{R}^n$ sufficiently small, $h(\delta_1, \ldots, \delta_n, z)$ vanishes for at least N distinct values of z.

At the same time, since $V(h)$ has zero Lebesgue measure (or, alternatively, by Schwartz–Zippel), there are values $\bar{\delta}_1, \ldots, \bar{\delta}_n \in (-\delta, \delta)$ and $\bar{z} \in \mathbb{R}$ with $h(\bar{\delta}_1, \ldots, \bar{\delta}_n, \bar{z}) \neq 0$. It follows that $h(\bar{\delta}_1, \ldots, \bar{\delta}_n, z)$ is a nonzero polynomial in z, of degree at most D by (C2), and hence $N \le D$ as claimed. It remains to prove the lemma.

Proof of Lemma 9.3. We will look for h in the form

$$h(y_1, \ldots, y_n, z) = \sum_{\alpha \in A} c_\alpha y_1^{\alpha_1} \cdots y_n^{\alpha_n} z^{\alpha_{n+1}},$$

where $A \subset \mathbb{Z}_{\ge 0}^{n+1}$ is a suitable finite set of $(n+1)$-tuples, whose choice we will discuss later, and where the c_α are regarded as unknowns. So we have $|A|$ unknowns.

If we make the substitution $y_1 = f_1, \ldots, y_n = f_n$, $z = \pi$ for a monomial $y_1^{\alpha_1} \cdots y_n^{\alpha_n} z^{\alpha_{n+1}}$, the degree of the resulting polynomial in $x_1, \ldots, x_n$ is

$$d_1\alpha_1 + \cdots + d_n\alpha_n + \alpha_{n+1}.$$

Let us call this expression the *weight* $w(\alpha)$, and set

$$w(A) := \max_{\alpha \in A} w(\alpha).$$

Thus, if we fix A, the degree of $\tilde{h}$, the polynomial after the substitution, is at most $w(A)$. Moreover, the coefficients of $\tilde{h}$ are *linear* functions of the c_α.

We want to force a zero coefficient for every monomial that could possibly appear in $\tilde{h}$; each such requirement yields a linear equation for the c_α. Since $\deg \tilde{h} \le w(A)$, we thus obtain $\binom{w(A)+n}{n}$ homogeneous linear equations for $|A|$ unknowns.

Hence the lemma will be proved as soon as we find A such that $|A| > \binom{w(A)+n}{n}$ and $\alpha_{n+1} \le D$ for all $\alpha \in A$.

For an integer W, let

$$A = A(W) := \{\alpha : w(\alpha) \le W, \alpha_{n+1} \le D\}.$$

We want to show that $|A(W)| > \binom{W+n}{n}$ holds for all sufficiently large W. The counting must be quite precise; after all, the proof cannot work with $D-1$ instead of D.

For a parameter T, let $N(T)$ be the number vectors $(\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n_{\ge 0}$ such that $\sum_{i=1}^n d_i \alpha_i \le T$; we have

$$|A(W)| = \sum_{\alpha_{n+1}=0}^{D} N(W - \alpha_{n+1}) \ge (D+1)N(W-D).$$

Let $B = B(T)$ be the set of all $\beta = (\beta_1, \ldots, \beta_n) \in \mathbb{Z}^n_{\ge 0}$ with $\beta_1 + \cdots + \beta_n \le T$; we have $|B| = \binom{T+n}{n}$. We can express $N(T)$ as the number of $\beta \in B$ such that $\beta_i \bmod d_i = 0$ for all i.

Let $r(\beta) = (\beta_1 \bmod d_1, \ldots, \beta_n \bmod d_n)$, and let us partition B into equivalence classes according to the value of $r(\beta)$; there are $d_1 d_2 \cdots d_n = D$ classes. It is easy to see that the class with $r(\beta) = (0, \ldots, 0)$ is at least as large as any other class, and so

$$N(T) \ge \frac{1}{D}\binom{T+n}{n}.$$

Consequently,

$$\begin{aligned}
|A(W)| &\ge (D+1)N(W-D) \ge \frac{D+1}{D}\binom{W-D+n}{n} \\
&= \frac{D+1}{D}\binom{W+n}{n} \cdot \frac{(W-D+n)\cdots(W-D+1)}{(W+n)\cdots(W+1)} \\
&\ge \binom{W+n}{n}\frac{D+1}{D}\left(1 - \frac{D}{W+1}\right)^n.
\end{aligned}$$

For D fixed and $W \to \infty$, we have $(1 - \frac{D}{W+1})^n \to 1$, while $\frac{D+1}{D}$ remains bounded away from 1. Hence $|A(W)| > \binom{W+n}{n}$ for W sufficiently large as desired. The lemma, as well as Bézout's inequality for nonsingular zeros, are proved. $\square$

10. Bounding the Number of Connected Components

How complicated can the zero set of a polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$ of degree d be? The answer depends, of course, on how we measure the complexity, and there are several sensible ways.

We will first look at the number of connected components of the complement, i.e., of $\mathbb{R}^n \setminus V(f)$. In this case a good answer can be given with a reasonably simple proof.

To see what can be expected, we consider the usual example with hyperplanes, slightly modified:

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} \prod_{j=1}^{m} (x_i - j).$$

The degree is $d = mn$, and the zero set, a grid of hyperplanes, partitions $\mathbb{R}^n$ into $(m+1)^n \sim (d/n)^n$ components (axis-parallel boxes).

Theorem 10.1. *For a polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$ of degree $d \geq 2$, $\mathbb{R}^n \setminus V(f)$ has at most $(d+1)^n$ components.*

The proof below is similar to one in [**ST12**, Appendix A]. This kind of arguments goes back to Oleinik and Petrovskiĭ [**OP49, Ole51**], Milnor [**Mil64**], and Thom [**Tho65**].

In the proof, we will need the following result.

Fact 10.2. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a polynomial map (that is, a map for which each coordinate $f_i : \mathbb{R}^n \to \mathbb{R}$ is given by a polynomial; in algebraic geometry, one usually speaks of* regular maps *in this context), and let $X \subset \mathbb{R}^n$ be a proper algebraic variety (that is, X is contained in the zero set of a nonzero polynomial). Then the image $f(X)$ does not fill any open ball in $\mathbb{R}^n$.*

This result may look obvious, but obvious approaches to proofs have their caveats.

First, we know that X is "small"; e.g., it does not fill any open ball. But, for example, the image of a segment under a continuous map may be a unit square, as is witnessed by the famous *Peano curve*. So we have to use other properties of f besides continuity.

Approaching from the side of mathematical analysis, we can use the fact (which we do not prove here) that the image Lebesgue null set under a smooth map is Lebesgue null, plus Exercise 2.3. In our case, a polynomial map is not only smooth (inifinitely differentiable), but also *locally Lipschitz*, which allows for a quite straightforward proof.

Exercise 10.3. (a) Verify that a polynomial map $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz, meaning that for every $x_0 \in \mathbb{R}^n$ there exist $\varepsilon > 0$ and L such that f is L-Lipschitz on the ε-ball around x_0, i.e., $\|f(x) - f(y)\| \leq L\|x - y\|$ for every choice of x, y in that ball. (Unlike in most uses of the letter ε in analysis, here one can actually take ε as large as desired.)

(b) Prove that the image of a Lebesgue null set under a locally Lipschitz map $\mathbb{R}^n \to \mathbb{R}^n$ is Lebesgue null. ✠

A more algebraic approach to Fact 10.2 is to prove that the image of a proper subvariety in $\mathbb{R}^n$ under a polynomial map is a proper subvariety of $\mathbb{R}^n$. Unfortunately, this is not literally true, as can be seen by modifying the hyperbola example from Section 8.2. What can be shown is that such an image is *contained* in a proper subvariety of $\mathbb{R}^n$, which is enough for our purposes. This is not too hard, given the tools covered so far, and it is a special case of a result mentioned in Section 8.3, stating that a regular map cannot increase dimension, but here we will not go through the argument.

Proof of Theorem 10.1. First we count only the *bounded* components of $\mathbb{R}^n \setminus V(f)$.

We do not know, a priori, that there are only finitely many components, but for some of the arguments below it will be important that we work with finitely many. So we fix any collection $\mathcal{C}$ of finitely many bounded components of $\mathbb{R}^n \setminus V(f)$ and work only with these. We will show that $|\mathcal{C}| \leq (d-1)^n$, which will imply, in particular, that there are only finitely many components altogether.

For each component $C \in \mathcal{C}$, we have either $f > 0$ or $f < 0$ on C; let us assume the former. We claim that f attains at least one maximum on C. Indeed, f attains some positive value $\varepsilon > 0$ at some point of C, the set $\{x \in C : f(x) \geq \frac{\varepsilon}{2}\}$ is compact and nonempty, and so f attains a maximum at some x_C there.

Since x_C lies inside the open set C and f is differentiable, the gradient ∇f vanishes at x_C, and hence $x_C \in V(\nabla f)$, where $V(\nabla f)$ is a shorthand for the set of common zeros of $\frac{\partial f}{\partial x_i}$, $i = 1, 2, \ldots, n$.

We note that $\deg \frac{\partial f}{\partial x_i} \leq d - 1$. The idea is to apply Bézout's inequality, in the form with nonsingular zeros, to bound $|V(\nabla f)|$, and hence the number of bounded components, by $(d-1)^n$.

The condition for nonsingularity of a common zero a of the $\frac{\partial f}{\partial x_i}$ reads $\det H_f(a) \neq 0$, where H_f is the *Hessian matrix* of f, with

$$(H_f)_{ij} := \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

However, we cannot guarantee that $\det H_f$ is not identically 0 (even some of the partial derivatives may be identically 0—for example, if f does not depend on some of the variables).

The next trick is to perturb the function whose maxima we seek. Indeed, if the maximum of f over a bounded component C is at least ε, then another function $\tilde{f}$ differing from f by at most $\frac{\varepsilon}{3}$, say, also has to attain a maximum in C. (Note that $C \in \mathcal{C}$ is still one of the original components of $\mathbb{R}^n \setminus V(f)$, even though we maximize the perturbed function $\tilde{f}$ over it.)

We actually make two perturbations. First, for δ sufficiently small, we set $\tilde{f} := f - \delta(x_1^2 + \cdots + x_n^2)$. This is the simplest kind of perturbation that may make the Hessian determinant nonzero (if we were willing to use Theorem 9.1 instead of Theorem 9.2, we could skip this perturbation).

It is easy to see that $H_{\tilde{f}} = H_f - 2\delta I$, where I is the identity matrix, and hence $\det H_{\tilde{f}} = 0$ exactly if 2δ is an eigenvalue of H_f. Thus, no matter what H_f looks like, $\det H_{\tilde{f}}$ is a nonzero polynomial for all but finitely many δ. We fix some sufficiently small δ for which $\det H_{\tilde{f}}$ is not identically zero; then $\tilde{f}$ is fixed too.

Next, we let $\tilde{f}_\eta := \tilde{f} - \eta_1 x_1 - \cdots - \eta_n x_n$, where $\eta = (\eta_1, \ldots, \eta_n)$ is a vector of parameters. Then $\nabla \tilde{f}_\eta = \nabla \tilde{f} - \eta$, and so instead of

counting the points in $V(\nabla\tilde{f}) = (\nabla\tilde{f})^{-1}(0)$, we now need to count the number of preimages of η under the polynomial map $\nabla\tilde{f}\colon \mathbb{R}^n \to \mathbb{R}^n$. (Geometrically, replacing $\tilde{f}$ with $\tilde{f}_\eta$ corresponds to slightly tilting the originally vertical direction in which we seek maxima or minima of $\tilde{f}$.)

We want to choose η sufficiently small (so that $\tilde{f}_\eta$ and f are sufficiently close) and such that the Hessian determinant $\det H_{\tilde{f}_\eta} = \det H_{\tilde{f}}$ does not vanish at the points of the preimage $(\nabla\tilde{f})^{-1}(\eta)$.

The variety of the Hessian determinant, $Y := V(\det H_{\tilde{f}})$, is the zero set of a nonzero polynomial, and $\nabla\tilde{f}$ is a polynomial mapping $\mathbb{R}^n \to \mathbb{R}^n$. Hence by Fact 10.2, there are arbitrarily small η avoiding the image of $\nabla\tilde{f}(Y)$.

For such η, all the maxima and minima of $\tilde{f}_\eta$ are nonsingular common zeros of the polynomials in $\nabla\tilde{f}_\eta$, and so we can bound their number by $(d-1)^n$ as desired.

It remains to account for the unbounded components. For that, we replace f with $g := f\cdot(x_1^2+\cdots+x_n^n-R^2)$, where R is a sufficiently large number; that is, to the zero set of f we add a large sphere. Then every component of $\mathbb{R}^n \setminus V(f)$ appears as a *bounded* component of $\mathbb{R}^n \setminus V(g)$. Since $\deg g = \deg f + 2$, the bound claimed in the theorem follows. □

A stronger version. The theorem just proved can be strengthened in several respects.

First, there is a quantitative improvement, which becomes significant if the degree d and the dimension n are comparable: the true bound is more like $(d/n)^n$ (which is the lower bound we got from the simple example) than d^n.

Second, the bound can be extended to the complement of the union of several zero sets, i.e., $\mathbb{R}^n \setminus (V(f_1) \cup \cdots \cup V(f_m))$. In this case a reasonably good bound can be obtained by simply setting $f = f_1 f_2 \cdots f_m$ and using the bound for a single polynomial.

Third, instead of considering just the complement, which is the set where all of $f_1,\ldots,f_m$ are nonzero, we can consider sets where

some of the f_i are required to be 0, some others positive, and some negative. These three improvements are all reflected in the next theorem.

Theorem 10.4. *Let $f_1, \dots, f_m \in \mathbb{R}[x_1, \dots, x_n]$ be polynomials of degree at most d, and for every sign vector $\sigma \in \{-1, 0, +1\}^m$ let $S_\sigma \subseteq \mathbb{R}^n$ be defined as*

$$\Big\{x \in \mathbb{R}^n : \operatorname{sgn} f_i(x) = \sigma_i \text{ for all } i = 1, 2, \dots, m\Big\}.$$

Then, for $m \geq n \geq 2$,

$$\sum_{\sigma \in \{-1,0,+1\}^m} \#S_\sigma \leq \left(\frac{50dm}{n}\right)^n,$$

where $\#S_\sigma$ denotes the number of connected components of S_σ.

The basic ideas of the proof are similar to those in the proof of Theorem 10.1 shown above, but the details are considerably more involved.

In the literature, such results are often stated as bounding the total topological complexity of the considered sets (more precisely, the sum of the *Betti numbers*) instead of just the number of connected components. For still other strengthenings of the just stated theorem, such as a more refined dependence on the degrees of the f_i, as well as replacing the ground set $\mathbb{R}^n$ with a k-dimensional algebraic variety in $\mathbb{R}^n$, see [**Bar13**] and references therein.

Bounds on the radius of components and inscribed balls. Another way of measuring zero sets of polynomials in $\mathbb{R}^n$ is, for example, by the radius of the smallest ball intersecting all connected components. Here, of course, we need to make some assumptions on the coefficients of the polynomials; typically we assume them to be integers not exceeding some given bound. Here is a general result of this kind:

Theorem 10.5. *Let $f_1, \dots, f_m \in \mathbb{Z}[x_1, \dots, x_n]$ be polynomials of maximum degree d whose coefficients are integers bounded by M in absolute value. For $\sigma \in \{-1, 0, 1\}^m$, let $S_\sigma := \{x \in \mathbb{R}^n : \operatorname{sgn} f_i(x) =$*

σ_i for all $i = 1, 2, \ldots, m\}$. Then each connected component of S_σ intersects the ball of radius $R = M^{(d+1)^{Cn}}$ centered at 0, where C is a suitable absolute constant. The bounded connected components of S_σ are all contained in that ball.

If $\sigma_i \neq 0$ for all i, or in other words, S_σ is defined only by strict inequalities, and if S_σ is nonempty, then it contains a rational point with coordinates whose numerators and denominators are integers not exceeding R in absolute value.

This kind of result goes back to [**GV88**, Lemma 9] (which deals with more special sets, namely, the zero set of a single polynomial), and the result, as above, about a ball intersecting all connected components is [**BPR96**, Theorem 4.1.1] (also see [**BPR03**, Theorem 13.14]). A statement directly implying the part with the ball containing all bounded components is [**BV07**, Theorem 6.2]. For the part with a rational point, see [**BPR03**, Theorem 13.15].

On applications. Theorem 10.4 and its relatives have probably hundreds of applications in geometry, combinatorics, computer science, and elsewhere. An old but still very beautiful one is Ben Or's lower bound method for algorithms described as algebraic computation trees [**BO83**].

Here is a quick application from [**AFR85**] which uses the more precise bound in Theorem 10.4. Let the sign pattern of an $n \times n$ matrix A be the matrix S with $s_{ij} = \operatorname{sgn} a_{ij}$. We claim that there are $n \times n$ matrices S with only ± 1 entries such that every A with sign pattern S has rank at least cn, for a positive constant c.

On the one hand, there are 2^{n^2} possible S's. On the other hand, an A of rank at most r can be written as UV^T, where U and V are $n \times r$ matrices. We consider the $2nr$ entries of U and V as variables; then the signs of the entries of A are signs of quadratic polynomials in these variables. We have $m = n^2$ polynomials and thus, by Theorem 10.4, there are no more than $O(n^2/nr)^{2nr}$ possible sign patterns of a rank-r matrix A. For $r < cn$ and c small, this quantity is smaller than 2^{n^2}, and so some patterns force rank at least cn.

11. Literature

Textbooks and lecture notes for such a classical subject as algebraic geometry abound, of course, but not all of them are equally accessible to beginners.

The usual hands-on introduction, with emphasis on computational aspects, is Cox, Little, and O'Shea [**CLO07**]. Schenck's book [**Sch03**] is very clear, readable, and concise; another advantage is that it also treats many related concepts from algebra and topology. A very good set of lecture notes freely accessible on the web, including some of the more advanced concepts, such as sheaves and schemes, is Gathmann [**Gat13**].

For intersection theory, dealing with generalizations of Bézout's theorem and other counting questions for varieties, a remarkable little book is Katz [**Kat06**], and an older concise introduction is Fulton [**Ful84**].

For combinatorial, geometric, and computer science applications of polynomials, we can recommend, for example, Chen, Kayal, and Wigderson [**CKW11**]. Recent treatments of methods similar to the one used in the joints problem are Guth [**Gut13**] and Tao [**Tao13**].

Acknowledgment. Many thanks to Emo Welzl for teaching, in his usual brilliant way, a course based on this chapter at the ETH together with J. M., and for many thoughtful comments. We would also like to thank Boris Bukh, Vincent Kusters, Zuzana Safernová, Adam Sheffer, and Noam Solomon for valuable comments, suggestions, and corrections to earlier versions of this chapter.

Bibliography

[AFR85] N. Alon, P. Frankl, and V. Rödl. Geometrical realization of set systems and probabilistic communication complexity. In *Proc. 26th IEEE Symposium on Foundations of Computer Science*, pages 277–280, 1985.

[Arr06] E. Arrondo. Another elementary proof of the Nullstellensatz. *Amer. Math. Monthly*, 113(2):169–171, 2006.

[Bar13] S. Barone. Some quantitative results in real algebraic geometry. Preprint, arXiv:1307.8353, 2013.

[BO83] M. Ben-Or. Lower bounds for algebraic computation trees. In *Proc. 15th Annu. ACM Sympos. Theory Comput.*, pages 80–86, 1983.

[BPR96] S. Basu, R. Pollack, and M.-F. Roy. On the combinatorial and algebraic complexity of quantifier elimination. *J. ACM*, 43(6):1002–1045, 1996.

[BPR03] S. Basu, R. Pollack, and M.-F. Roy. *Algorithms in Real Algebraic Geometry*. Algorithms and Computation in Mathematics 10. Springer, Berlin, 2003.

[BV07] S. Basu and N. N. Vorobjov. On the number of homotopy types of fibres of a definable map. *J. Lond. Math. Soc., II. Ser.*, 76(3):757–776, 2007.

[CKW11] Xi Chen, N. Kayal, and A. Wigderson. Partial derivatives in arithmetic complexity and beyond. *Found. Trends Theor. Comput. Sci.*, 6(1-2):1–138, 2011.

[CLO05] D. A. Cox, J. Little, and D. O'Shea. *Using Algebraic Geometry*. Springer, New York, 2005.

[CLO07] D. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms*. Undergraduate Texts in Mathematics. Springer, New York, third edition, 2007.

[Ful84] W. Fulton. *Introduction to Intersection Theory in Algebraic Geometry*, volume 54 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1984.

[Gal14] F. Le Gall. Powers of tensors and fast matrix multiplication. Preprint arXiv:1401.7714, 2014.

[Gat13] A. Gathmann. Algebraic geometry. Lecture Notes, TU Kaiserslautern, `http://www.mathematik.uni-kl.de/agag/mitglieder/professoren/gathmann/notes/alggeom/`, 2013.

[GK10] L. Guth and N. H. Katz. Algebraic methods in discrete analogs of the Kakeya problem. *Adv. Math.*, 225(5):2828–2839, 2010.

[Gut13] L. Guth. The Polynomial Method, 2013. Book in preparation.

[GV88] D. Yu. Grigor'ev and N. N. Vorobjov jun. Solving systems of polynomial inequalities in subexponential time. *J. Symb. Comput.*, 5(1-2):37–64, 1988.

[Har09] N. J. A. Harvey. Algebraic algorithms for matching and matroid problems. *SIAM J. Comput.*, 39(2):679–702, 2009.

[Hei83] J. Heintz. Definability and fast quantifier elimination in algebraically closed fields. *Theoret. Comput. Sci.*, 24(3):239–277, 1983. Corrigendum ibid. 39,1983: 2–3.

[Kat06] S. Katz. *Enumerative Geometry and String Theory*, volume 32 of *Student Mathematical Library*. American Mathematical Society, Providence, RI, 2006. IAS/Park City Mathematical Subseries.

[Kol88] J. Kollár. Sharp effective Nullstellensatz. *J. Amer. Math. Soc.*, 1(4):963–975, 1988.

[Mil64] J. W. Milnor. On the Betti numbers of real algebraic varieties. *Proc. Amer. Math. Soc.*, 15:275–280, 1964.

[MO02] C. Michaux and A. Ozturk. Quantifier elimination following Muchnik. Univ. de Mons-Hainaut Preprint Series (#10), 2002.

[Ole51] O. A. Oleinik. Estimates of the Betti numbers of real algebraic hypersurfaces (in Russian). *Mat. Sbornik (N. S.)*, 28(70):635–640, 1951.

[OP49] O. A. Oleinik and I. B. Petrovskiĭ. On the topology of of real algebraic surfaces (in Russian). *Izv. Akad. Nauk SSSR*, 13:389–402, 1949.

[Sch95] J. Schmid. On the affine Bézout inequality. *Manuscripta Math.*, 88(2):225–232, 1995.

[Sch03] H. Schenck. *Computational Algebraic Geometry*, volume 58 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2003.

[ST12] J. Solymosi and T. Tao. An incidence theorem in higher dimensions. *Discrete Comput. Geom.*, 48(2):255–280, 2012.

[Tao12] T. Tao. Spending Symmetry. Book in preparation, draft available at `http://terrytao.wordpress.com/books/spending-symmetry/`, 2012.

[Tao13] T. Tao. Algebraic combinatorial geometry: the polynomial method in arithmetic combinatorics, incidence combinatorics, and number theory. Preprint, arXiv:1310.6482, 2013.

[Tho65] R. Thom. On the homology of real algebraic varieties (in French). In S. S. Cairns, editor, *Differential and Combinatorial Topology*. Princeton Univ. Press, 1965.

[Wal79] M. Waldschmidt. *Transcendence Methods*. Queen's University, 1979. Available at `http://www.math.jussieu.fr/~miw/articles/pdf/QueensPaper52.pdf`.

[Woo96] T. D. Wooley. A note on simultaneous congruences. *J. Number Theory*, 58(2):288–297, 1996.

Chapter 6

Topology

Topology has spectacular applications in discrete mathematics and computer science, such as in lower bounds for the chromatic number of graphs (which will be discussed later to some extent), in results about the behavior of distributed computing systems (see Herlihy, Kozlov, and Rajsbaum [**HKR13**]), or in methods for reconstructing 3-dimensional shapes from point samples, whose importance increases with the advent of ubiquitous 3D printing. Among applied fields using topology we mention computational biology and medical informatics.

Yet, in the authors' experience, the entrance barriers of topology are relatively high. This has to do with the extent, maturity, and technical sophistication of the field. At the very beginning of serious study, a newcomer is confronted with new language and conventions, such as commutative diagrams, exact sequences, and concepts of category theory. At the same time, in order to honestly reach the first real results, one also has to work through a number of technicalities such as approximations of continuous maps. These things can be experienced once and then more or less forgotten, yet skipped they should not be. Last, but not least, some of the fundamental concepts are truly sophisticated.

The notion of homology seems to be a particularly high stumbling block. Many people with some topological background switch off when a homology or cohomology group appears on the board. In

this chapter, we aim at an introduction with as few technicalities as possible reaching all the way to (simplicial) homology groups, including the fact that they are independent of the choice of triangulation. The latter is technical, but we do not see any other way of getting used to the machinery without actually working through a number of the details.

The chapter does not get one very far in topology, but it may make a systematic study of full-fledged textbooks easier for those wishing to explore more deeply.

We fix notation for two sets in $\mathbb{R}^n$, which are used all the time in topology. The n-dimensional **ball** is

$$B^n = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$$

(some sources prefer the word *disk* and the notation D^n), and the $(n-1)$-dimensional **sphere** is the boundary of B^n, i.e.,

$$S^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$$

(note that S^2 lives in $\mathbb{R}^3$). Both are considered with the Euclidean metric.

1. Topological Spaces and Continuous Maps

A topological space is a mathematical structure for capturing the notion of continuity, one of the most basic concepts of all mathematics, on a very general level.

The usual definition of continuity of a map from introductory courses uses the notion of distance: a map is continuous if the images of sufficiently close points are again close.

This can be formalized for maps between metric spaces. We recall that a **metric space** is a pair (X, d_X), where X is a set and $d_X\colon X \times X \to \mathbb{R}$ is a **metric** satisfying several natural axioms (x, y, z are arbitrary points of X): $d_X(x,y) \geq 0$, $d_X(x,x) = 0$, $d_X(x,y) > 0$ for $x \neq y$, $d_X(y,x) = d_X(x,y)$, and $d_X(x,y) + d_X(y,z) \geq d_X(x,z)$ (the *triangle inequality*). The most important example of a metric space is $\mathbb{R}^n$ with the Euclidean metric, and another, of particular interest in computer science, is a graph with the shortest-path metric.

Formally, a map $f\colon X \to Y$ between metric spaces is continuous if for every $x \in X$ and every $\varepsilon > 0$ there exists $\delta > 0$ such that whenever $y \in X$ and $d_X(x,y) < \delta$, we have $d_Y(f(x), f(y)) < \varepsilon$.

One can think of a topological space as starting with a metric space and forgetting the metric, remembering only which sets are open. (We recall that a set $U \subseteq X$ in a metric space is open if for every $x \in U$ there is $\varepsilon > 0$ such that U contains the ε-ball around x.) This is not quite precise since topological spaces are much more general than metric spaces and there are many interesting specimens which cannot be obtained from any metric space, but in applications of topology we mostly encounter topological spaces coming from metric ones.

Topological space. Here is the general definition.

Definition 1.1. *A* **topological space** *is a pair $(X, \mathcal{O})$, where X is a ground set and $\mathcal{O} \subseteq 2^X$ is a set system, whose members are called the* **open sets***, such that $\emptyset \in \mathcal{O}$, $X \in \mathcal{O}$, the intersection of finitely many open sets is an open set, and so is the union of an arbitrary collection of open sets.*

The system $\mathcal{O}$ as in the definition is sometimes called a **topology** on X.

In this chapter, we will often say just **space** instead of topological space.

Two topological spaces $(X, \mathcal{O}_X)$ and $(Y, \mathcal{O}_Y)$ are considered "the same" from the point of view of topology if there is a bijective map $f\colon X \to Y$ that preserves open sets in both directions; that is, $V \in \mathcal{O}_Y$ implies $f^{-1}(V) \in \mathcal{O}_X$ and $U \in \mathcal{O}_X$ implies $f(U) \in \mathcal{O}_Y$. For most mathematical structures, such as groups or graphs, an f with analogous structure-preserving properties is called an isomorphism, but in topology an f as above is called a **homeomorphism**. Topological spaces X and Y are said to be **homeomorphic**, written $X \cong Y$, if there is a homeomorphism between them. (Strictly speaking, we should write that the topological spaces $(X, \mathcal{O}_X)$ and $(Y, \mathcal{O}_Y)$ are homeomorphic, but in agreement with a common practice we mostly use the same letter for the topological space and for the underlying set.)

Here we see a substantial difference between metric and topological spaces: two spaces which are metrically quite different can be homeomorphic and thus topologically the same.

Exercise 1.2. Verify that the following pairs of spaces are homeomorphic (the topology is always given by the Euclidean metric):

(a) $\mathbb{R}$, the open interval $(0,1)$, and $S^1 \setminus \{(0,1)\}$ (the unit circle in the plane minus one point).

(b) S^1 and the boundary of the unit square $[0,1]^2$. ✠

Similarly, different metrics on X may induce the same topology: this is the case for all ℓ_p metrics on $\mathbb{R}^n$ (n fixed), for example. For readers familiar with Banach spaces we also mention that all infinite-dimensional separable Banach spaces are homeomorphic as topological spaces—this is a nontrivial theorem of Kadets; in this case, from the point of view of functional analysis, the topology carries too little information.

Subspaces. The topological spaces encountered most often in applications, as well as in a substantial part of topology itself, are subspaces of some $\mathbb{R}^n$ with the standard topology (i.e., the one induced by the Euclidean metric), or are at least homeomorphic to such subspaces.

In general, for a topological space $(X, \mathcal{O})$, every subset $Y \subseteq X$ induces a **subspace** of $(X, \mathcal{O})$, namely, the topological space $(Y, \{U \cap Y : U \in \mathcal{O}\})$. (This is quite different, e.g., from groups, where only quite special subsets correspond to subgroups.) Note that the open sets of the subspace need not be open as subsets of X: for instance, let X be the Euclidean plane and Y a segment in it; then Y is open in Y but, of course, not in the plane.

Neighborhoods, bases, closure, boundary, interior. A set N in a topological space X is called a **neighborhood** of a point $x \in X$ if there is an open set U such that $x \in U \subseteq N$.

The system $\mathcal{O}$ of all open sets in a topological space can often be described more economically by specifying a **base** of $\mathcal{O}$, which is a collection $\mathcal{B} \subseteq \mathcal{O}$ such that every $U \in \mathcal{O}$ is a union of some of the sets in $\mathcal{B}$. For example, the system of all open intervals is a base of

the standard topology of $\mathbb{R}$, and so is the system of all open intervals with rational endpoints.

Exercise 1.3. Check that the system of all open balls of radius $\frac{1}{n}$, $n = 1, 2, \ldots$, constitutes a base of the topology of a metric space. ✠

A possibly still more compact specification of a topology $\mathcal{O}$ is a *subbase*, which is a system $\mathcal{S}$ such that the system of all finite intersections of sets from $\mathcal{S}$ forms a base of $\mathcal{O}$. An example is the system of all intervals $(-\infty, a)$ and (a, ∞), $a \in \mathbb{R}$, for $\mathbb{R}$.

A set $F \subseteq X$ is **closed** if $X \setminus F$ is open. Traditionally, one uses letters U, V, W for open sets and F, G, H for closed sets, and in sketches, open sets are drawn as smooth ovals and closed sets as polygons.

The **closure** $\mathrm{cl}\, Y$ of a set in a topological space X is the intersection of all closed sets containing Y (an alternative notation is $\overline{Y}$). In the metric case, the closure consists of all points with zero distance to Y (where $d_X(x, Y) = \inf_{y \in Y} d_X(x, y)$). The **boundary** of Y is $\partial Y := \mathrm{cl}(Y) \cap \mathrm{cl}(X \setminus Y)$, and the **interior** $\mathrm{int}\, Y$ consists of all $x \in Y$ that have an open neigborhood contained in Y.

We note that these last three notions depend not only on Y, but also on the space X in which they are considered: for example, if $X = \mathbb{R}$ and Y is the closed interval $[0, 1]$, then $\partial Y = \{0, 1\}$ and $\mathrm{int}\, Y = (0, 1)$, but if we consider the segment Y' connecting the points $(0, 0)$ and $(1, 0)$ as a subspace of $\mathbb{R}^2$, then $Y' \cong Y$ but $\partial Y' = Y'$ and $\mathrm{int}\, Y' = \emptyset$. To avoid ambiguities one sometimes writes $\mathrm{cl}_X\, Y$, $\partial_X Y$, $\mathrm{int}_X\, Y$.

Continuous maps. Now we return to continuity, whose topological definition is strikingly simple.

Definition 1.4. *A* **continuous map** *of a topological space* $(X, \mathcal{O}_X)$ *into a topological space* $(Y, \mathcal{O}_Y)$ *is a map* $f\colon X \to Y$ *of the underlying sets such that* $f^{-1}(U) \in \mathcal{O}_X$ *for all* $U \in \mathcal{O}_Y$*. In words, a map is continuous if the preimages of all open sets are open.*

In topological texts, *all maps between topological spaces are usually assumed to be continuous* unless stated otherwise. We will also sometimes use this convention.

The next exercise is definitely worth doing.

Exercise 1.5. Show that for maps $\mathbb{R} \to \mathbb{R}$ (where $\mathbb{R}$ has the standard topology), or more generally for maps between metric spaces, Definition 1.4 of continuity is equivalent to the ε-δ definition recalled earlier. ✠

Exercise 1.6. A curious reader might ask why the definition of continuity requires preimages, rather than images, of open sets to be open. We define a map $f\colon X \to Y$ between topological spaces to be an *open map* if $f(U)$ is open for every open set U. Find examples, involving maps between subspaces of $\mathbb{R}$, of a continuous map that is not open, as well as of an open map that is not continuous. ✠

Exercise 1.7. (a) Check that a homeomorphism of topological spaces can equivalently be defined as a bijective continuous map with continuous inverse.

(b) Find an example of a bijective continuous map between suitable subspaces of $\mathbb{R}$ that is not a homeomorphism. ✠

Exercise 1.8. Let X, Y be a topological spaces, let $f\colon X \to Y$ be a map, and let $A_1, \ldots, A_n \subseteq X$ be closed sets that together cover all of X. Let us assume that the restriction of f to the subspace of X induced by A_i is continuous, for every $i = 1, 2, \ldots, n$ (while we do not a priori assume f continuous). Prove that f is continuous. ✠

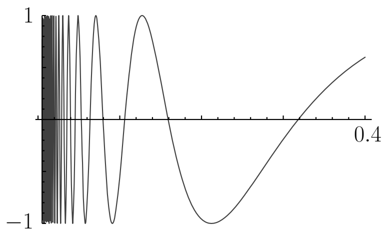
2. Bits of General Topology

There is a sizeable list of properties a topological space may or may not have. (These properties are all invariant under homeomorphism.) Here we present a brief selection.

Connectedness. There are two different definitions capturing the intuitive idea that a topological space “has just one piece.” A topological space X is **connected** if X cannot be written as a union of

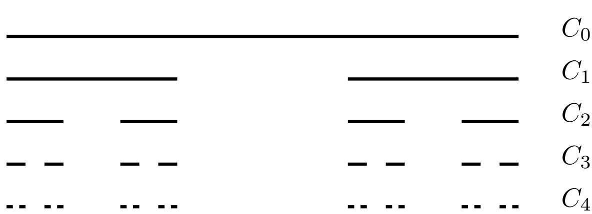
two disjoint nonempty open sets.[1] And X is **path-connected** if every two points x, y are connected by a path, where in the topological setting, a **path** from x to y is a continuous map $f\colon [0,1] \to X$ of the unit interval with $f(0) = x$ and $f(1) = y$.

Connectedness and path-connectedness are not equivalent: the latter implies the former, but a famous example of a connected space that is not path-connected is the *topologist's sine curve*, the subspace of $\mathbb{R}^2$ consisting of the vertical segment from $(0, -1)$ to $(0, 1)$ and the graph of the function $x \mapsto \sin \frac{1}{x}$ for $x > 0$:



For applications, path-connectedness seems to be more important.

One can define *connected components* of a space X as inclusion-maximal subsets that, considered as topological subspaces of X, are connected. Path-connected components can be defined analogously. Among the wilder examples we have the famous **Cantor set** $C \subset \mathbb{R}$, given by $C = \bigcap_{i=1}^{\infty} C_i$, where $C_0 = [0, 1]$ and $C_i = \frac{1}{3}C_{i-1} \cup (\frac{1}{3}C_{i-1} + \frac{2}{3})$:



All of its connected (or path-connected) components are singletons, and there are uncountably many.

[1] The literature is not quite unified concerning the question of whether the empty topological space is connected. It should be according to the general definition, but for many purposes it is better to define that it is not.

For every topological property we can hope that it allows us to distinguish some pairs of non-homeomorphic spaces. In the case of (path-)connectedness, we can prove that no two of the spaces S^1, $(0,1)$ (open interval) and $[0,1]$ (closed interval) are homeomorphic: indeed, if we remove a single point, then S^1 *always* stays connected, $(0,1)$ *never*, and $[0,1]$ *sometimes* stays connected and sometimes not. (Can you see other ways of proving any of these non-homeomorphisms?)

Bizarre spaces and general topology. So far we may have made the impression that all topological spaces look more or less like subspaces of Euclidean spaces, but this is very far from the truth—they need not even look like metric spaces.

A topological space X whose topology can be obtained from some metric is called **metrizable**. A traditional subfield of topology called **general topology**, or *point-set topology*, mainly studies various properties of topological spaces more general than metrizability, relations among them, conditions making a space metrizable, etc.

Let us list several examples taken from the vast supply built in general topology over the years. We will not prove any properties for them, except possibly in exercises—the intention is to give the reader some feeling for the possible pathologies occurring in arbitrary topological spaces, as well as a supply of candidate counterexamples for refuting too general claims. The reader may at least want to check in passing that these are indeed topological spaces.

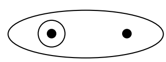
All of the examples except for (A) are non-metrizable, which in several cases is nontrivial to prove.

(A) Any set X, such as the real numbers, can be given the **discrete topology**, in which all subsets are open. Note that the integers $\mathbb{Z}$ inherit such a topology as a subspace of $\mathbb{R}$ with the standard topology, but discrete topology becomes more exotic if the ground set is uncountable.

(B) Let X be an infinite set. The *topology of finite complements* has open sets $\emptyset$ and $X \setminus B$ for all $B \subseteq X$ finite. Similarly one can define the topology of *countable complements* on an uncountable set.

(C) We recall that an *algebraic variety* in $\mathbb{R}^n$ (or, for that matter, in $\mathbb{K}^n$ for any field $\mathbb{K}$) is the set of common zeros of a set of n-variate polynomials. The open sets of the *Zariski topology* on $\mathbb{R}^n$ has all complements of algebraic varieties as open sets. The reader may want to check that for $n = 1$ we get the topology of finite complements. This is a (somewhat rare) example of an exotic topology used heavily outside the field of general topology, namely, in algebraic geometry.

(D) The two-point space $\{1, 2\}$ in which open sets are $\emptyset$, $\{1\}$, and $\{1, 2\}$:



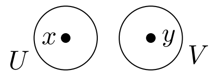
Here the closure of the singleton set $\{1\}$ is $\{1, 2\}$, while 1 is not in the closure of $\{2\}$, which probably cannot be considered good manners.

(E) The *Sorgenfrey line* is $\mathbb{R}$ with the topology whose base are all half-open intervals $[a, b)$. The *Sorgenfrey plane* is the product of the Sorgenfrey line with itself (products will be introduced soon); explicitly, this is $\mathbb{R}^2$ with the topology whose base are half-open rectangles $[a, b) \times [c, d)$.

(F) Let ω_1 be the first uncountable ordinal (assuming that the reader knows or looks up what ordinal numbers are). The set $L = \omega_1 \times [0, 1)$ is ordered lexicographically, and then given the topology whose base are all open intervals in this linear ordering. The resulting topological space is called the *long ray*; locally it looks like $\mathbb{R}$ with the standard topology, but globally it is "too long" to be metrizable.

Separation axioms. One class of properties intended to measure how close a given space is to metrizability are traditionally called the *separation axioms*. The most popular ones are called T_0, T_1, T_2, T_3, $T_{3\frac{1}{2}}$, T_4 in the order of increasing strength (T abbreviates the German *Trennungsaxiom*, i.e., separation axiom), and one can also find $T_{2\frac{1}{2}}$, T_5, and T_6 in the literature, plus a number of others not quite fitting the T_i scale. Metrizable spaces have all of these properties.

Probably the most important to remember is T_2: a space X is T_2 or **Hausdorff** if for every two distinct points $x, y \in X$ there are open sets $U \ni x$ and $V \ni y$ with $U \cap V = \emptyset$. Briefly, distinct points can be separated by open sets:



Decent topological spaces are at least Hausdorff (possibly with the honorable exception of the Zariski topology); examples (B)–(D) above are not.

For illustration, we also mention that a T_3 or *regular space* is one that is T_2 and in which every closed set F can be separated from every point $x \notin F$ by open sets, while a T_4 or *normal space* is T_2 and disjoint closed sets can be separated by open sets:



There are examples showing that all of the hierarchy is strict, i.e., T_i does not imply T_j for $i < j$. Sometimes these are quite sophisticated, the hardest being one showing $T_3 \not\Rightarrow T_{3\frac{1}{2}}$. As far as our examples above are concerned, the Sorgenfrey plane is $T_{3\frac{1}{2}}$ but not T_4.

We conclude this brief mention of the separation axioms by a warning: The literature is far from unified concerning terminology. The main difference is in whether, for the higher separation axioms like T_3 or T_4, one automatically assumes T_1 (or, equivalently, T_2) or not. Indeed, the modern usage seems to prefer "normal" to mean "disjoint closed sets separable by open sets" while T_4 means "normal+T_1." So it is advisable to check the definitions carefully.

Cardinality restrictions. A very important notion is that of a dense subset: a set $D \subseteq X$ is **dense** in a topological space X if $\operatorname{cl} D = X$.

A space X is **separable** if it has a countable dense set. The space $\mathbb{R}^n$ with the standard topology is separable because the set $\mathbb{Q}^n$ of all rational points is dense in it, and so is every subspace.

Exercise 2.1. (a) Show that the Sorgenfrey plane in (E) above is separable, but it has a non-separable subspace.

(b) Prove that a subspace of a separable metric space is separable.

✠

A notion with less importance outside topology is a space with countable base (meaning a base for its topology as introduced earlier), which for historical reasons is often called a *second-countable space*. This is a property much stronger than separability.

Polish spaces. In many fields of mathematics, when one wants to work only with "sufficiently nice" topological spaces, one makes assumptions even stronger than metrizability. The most frequent such concept is perhaps a **Polish space**, which is a separable completely metrizable space.

Here one needs to know that a *complete metric space* is one in which every Cauchy sequence[2] converges to a limit. For example, the Euclidean metric on $\mathbb{R}$ is complete, but on $(0, 1)$ it is not. The definition of Polish space requires the existence of at least one complete metric inducing the topology; so, for example, $(0, 1)$ is a Polish space.

Let us conclude this section with two examples of nice basic theorems of general topology. The first one we state without proof:

Theorem 2.2 (Tietze extension theorem)**.** *Let X be a metric space, or more generally, a T_4 topological space, let $A \subseteq X$ be closed, and let $f\colon A \to \mathbb{R}$ be a continuous map. Then there exists a continuous extension $\overline{f}\colon X \to \mathbb{R}$ of f, for which we may moreover assume* $\sup_{x\in X} |\overline{f}(x)| \le \sup_{a\in A} |f(x)|$.

Theorem 2.3 (Urysohn metrization theorem)**.** *Every T_3 topological space with a countable base is metrizable.*

---

[2] A sequence $(x_1, x_2, \ldots)$ is *Cauchy* if for every $\varepsilon > 0$ there is n such that for all $i, j \ge n$ we have $d_X(x_i, x_j) < \varepsilon$.

We present a proof, assuming for convenience T_4 instead of just T_3.

Exercise 2.4. Prove that every T_3 space with a countable base is also T_4. ✠

The proof of Theorem 2.3 contains a very useful and general trick (appearing frequently in various areas, such as theory on low-distortion embeddings of finite metric spaces, a recent hot topic in computer science).

The countable base assumption, as well as Tietze's extension theorem, are used in the next lemma.

Lemma 2.5. *For every T_4 space X with a countable base there exists a countable sequence $(f_1, f_2, \ldots)$ of continuous functions $X \to [0,1]$ such that for every point $x \in X$ and every open set U with $x \in U$ there is an f_i that is 0 outside U and 1 in x.*

Proof. For every pair (B, B') of the assumed countable base $\mathcal{B}$ of X with $\operatorname{cl} B' \subset B$, we use the Tietze extension theorem to get a function $X \to [0,1]$ that equals 1 on $\operatorname{cl} B'$ and equals 0 on $X \setminus B$. These are the desired f_i.

To check that this works, we consider $x \in U$ as in the lemma. We find $B \in \mathcal{B}$ with $x \in B \subseteq U$, and then we use the T_3 property to separate x from $X \setminus B$ by disjoint open sets $V \ni x$ and $W \supseteq X \setminus B$. It follows that $\operatorname{cl} V \subseteq X \setminus W \subseteq B$. Finally we shrink V to some $B' \in \mathcal{B}$ still containing x.

We now have $x \in B' \subseteq \operatorname{cl} B' \subseteq B \subseteq U$, and it is clear that the separating function made above for (B, B') is 1 at x and 0 outside U. □

Proof of Theorem 2.3 under the T_4 assumption. Let H, the **Hilbert cube**, be the metric space of all infinite sequences $x = (x_1, x_2, \ldots)$, $x_i \in [0, \frac{1}{i}]$, $i = 1, 2, \ldots$, with the ℓ_2 metric, meaning that the distance of x and y is $\left(\sum_{i=1}^{\infty}(x_i - y_i)^2\right)^{1/2}$.

We will show that the space X as in the theorem is homeomorphic to a subspace of H. Then the metrizability of X will be clear.

We define a map $f\colon X \to H$ by

$$f(x) := \left(\tfrac{1}{1} f_1(x), \tfrac{1}{2} f_2(x), \tfrac{1}{3} f_3(x), \ldots\right)$$

where the f_i are as in Lemma 2.5 (this definition is the main trick!).

Exercise 2.6. Check that f is continuous (this uses nothing but the continuity of the f_i) and injective. ✠

It remains to verify that the inverse map $f^{-1}\colon f(X) \to X$ is continuous. To this end it suffices to check that for every $U \subseteq X$ open and every $x \in U$, there is an $\varepsilon > 0$ such that $f(U)$ contains the ε-ball around $f(x)$ (ball in $f(X)$, not in all of H, that is).

As expected, we fix i with $f_i(x) = 1$ and f_i zero outside U, and we let $\varepsilon := \frac{1}{2i}$. Now we suppose that $y \in X$ is such that $f(x)$ and $f(y)$ have distance at most ε in H; we want to conclude $y \in U$. We have, in particular, $\frac{1}{i}|f_i(x) - f_i(y)| \le \varepsilon$, so $f_i(y) \ge \frac{1}{2}$, and thus $f_i(y) \ne 0$. Hence $y \in U$ as needed. □

3. Compactness

One of the most important and most applied topological properties is *compactness*. Intuitively, a compact space is one that does not have too much room inside. The topological definition is quite simple:

Definition 3.1. *A topological space X is* **compact** *if for every collection $\mathcal{U}$ of open sets in X whose union is all of X, there exists a finite $\mathcal{U}_0 \subseteq \mathcal{U}$ whose union also covers all of X. In brief, every open cover of X has a finite subcover.*

A set $C \subseteq X$ is a **compact set** in X if C with the subspace topology is a compact space.

The notion of compactness was first developed in the metric setting, with a different definition, which is still presented in many introductory courses. Namely, a metric space X is compact if every infinite sequence $(x_1, x_2, \ldots)$ contains a subsequence $(x_{i_1}, x_{i_2}, \ldots)$, $i_1 < i_2 < \cdots$, that is convergent.

Exercise 3.2. Prove that if X is a metric space that is compact according to Definition 3.1, then every infinite sequence has a convergent subsequence. Hint: construct an open cover by balls “witnessing” that there is no convergent subsequence. ✠

Diligent readers may also do the opposite implication for metric spaces, but this is more difficult.

While one can naturally define convergent sequences in a topological space, and thus transfer the definition with sequences to topological spaces, one obtains a different, and much less well behaved, notion of *sequential compactness*. From this point of view, the topological approach, as opposed to the metric one, greatly clarified the essence of the notion.

Mainly in order to show typical proofs in general topology, we will now develop some properties of compactness, culminating in two extremely useful results concerning compact sets.

Lemma 3.3.

(i) *A closed subset of a compact space is compact.*

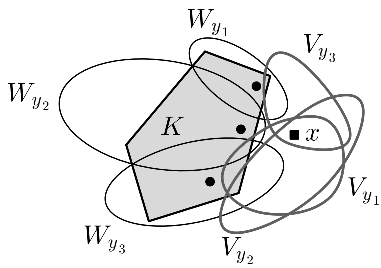
(ii) *A compact subset in a Hausdorff space is closed.*

(iii) *If $f\colon X \to Y$ is continuous and $K \subseteq X$ is compact, then $f(K)$ is compact (and hence closed if Y is Hausdorff).*

To appreciate (iii), one should realize that continuous maps need not map closed sets to closed sets in general.

Proof. In (i), let X be compact and $F \subseteq X$ be closed. Consider an open cover $\mathcal{U}$ of F, and for every $U \in \mathcal{U}$, fix an open set $\tilde{U}$ in X with $\tilde{U} \cap K = U$. Then $\tilde{\mathcal{U}} := \{\tilde{U} : U \in \mathcal{U}\} \cup \{X \setminus F\}$ is an open cover of X. From a finite subcover of $\tilde{\mathcal{U}}$ we obtain a finite subcover of $\mathcal{U}$ by restricting everything back to F.

For (ii), let X be Hausdorff and $K \subseteq X$ be compact. It suffices to show that for every $x \notin K$ there is an open U_x such that $U_x \cap K = \emptyset$. For every $y \in K$ we can fix, by the Hausdorff property, disjoint open sets $V_y \ni x$ and $W_y \ni y$. The W_y for all $y \in K$ form an open cover of K, so we select a finite subcover, say $W_{y_1}, \ldots, W_{y_n}$, and we set $U_x := \bigcap_{i=1}^{n} V_{y_i}$.



Finally, (iii) is easy based on the observation that if $\mathcal{U}$ is an open cover of $f(K)$, then $\{f^{-1}(U) : U \in \mathcal{U}\}$ is an open cover of K. □

Here is the first often-applied result.

Theorem 3.4. *Let K be compact, and let $f\colon K \to \mathbb{R}$ be a continuous function. Then f attains its minimum: there exists $x_0 \in K$ with $f(x_0) = \inf_{x\in K} f(x)$.*

In particular, a continuous function on a compact set is bounded, and a function on a compact set that is never zero is bounded away from 0; that is, there is $\varepsilon > 0$ such that $|f(x)| \geq \varepsilon$ for all $x \in K$.

Proof. By Lemma 3.3(iii), $Y := f(K) \subseteq \mathbb{R}$ is compact. Set $m := \inf Y$, choose a sequence $(y_1, y_2, \ldots)$, $y_i \in Y$, converging to m, and set $U_i := (y_i, \infty)$.

If the U_i do not cover Y, then this can be only because they all avoid m, and in particular, $m \in Y$. So we suppose that $\{U_i\}$ is an open cover of Y, and we select a finite subcover $U_{i_1}, \ldots, U_{i_n}$. Let $y^* := \min\{y_{i_1}, \ldots, y_{i_n}\}$. Then $Y \subseteq \bigcup_{j=1}^{n} U_{i_j} = (y^*, \infty)$, but this is a contradiction since $y^* \in Y$. □

Products. The product of two topological spaces $(X, \mathcal{O}_X)$ and $(Y, \mathcal{O}_Y)$ is defined in an expected way, with the ground set $X \times Y$ and the collection $\{U \times V : U \in \mathcal{O}_X, V \in \mathcal{O}_Y\}$ of *open rectangles* as a base of the topology.

The definition of a product of *infinitely many* spaces is trickier (but often needed): we do not take all open rectangles, but only those having only finitely many coordinates in which the open set is not the

whole space. Thus, if $(X_i, \mathcal{O}_i)_{i\in I}$ is a collection of spaces indexed by an arbitrarily large set I, then the product space $\prod_{i\in I}(X_i, \mathcal{O}_i)$ has ground set $\prod_{i\in I} X_i$, and a base of the topology is

$$\Big\{\prod_{i\in I} U_i : U_i \in \mathcal{O}_i, |\{i \in I : U_i \neq X_i\}| < \infty\Big\}.$$

For example, the product of countably many copies of the two-point discrete space $\{0, 1\}$ turns out to be homeomorphic to the Cantor set C, and the product of countably many copies of $\{0, 1, 2, \ldots\}$, again with the discrete topology, is homeomorphic to the set of all irrational numbers with the standard topology inherited from $\mathbb{R}$ (ambitious readers may want to prove these).

Exercise 3.5. Prove that a product of Hausdorff spaces is Hausdorff. ✠

Theorem 3.6 (Tychonoff's theorem)**.** *The product of an arbitrary collection of compact topological spaces is compact.*

Exercise 3.7. (a) Prove that if $X\times Y$ is a product of two topological spaces such that every open cover of $X \times Y$ by open rectangles (i.e., sets of the form $U \times V$, U open in X, V open in Y) has a finite subcover, then $X \times Y$ is compact.

(b) Prove Tychonoff's theorem for products of two spaces. ✠

The proof of Tychonoff's theorem for infinitely many factors needs more work, and more significantly, it relies on the *axiom of choice*—Tychonoff's theorem is actually one of the important theorems *equivalent* to the axiom of choice.

Instead of a proof, we will demonstrate a typical combinatorial application (similar considerations underlie compactness principles in logic and elsewhere). We recall that a graph $G = (V, E)$ is *k-colorable* if there is a map (coloring) $c\colon V \to [k] := \{1, 2, \ldots, k\}$ such that $f(u) \neq f(v)$ whenever $\{u, v\}$ is an edge of G.

Proposition 3.8. *Let G be an infinite graph. If every finite subgraph of G is k-colorable, then G is k-colorable.*

For countable graphs there is an elementary inductive proof. Tychonoff's theorem provides a quick proof in general.

Proof. For every vertex $v \in V$, let X_v be a copy of the discrete topological space $[k]$, and let $X := \prod_{v \in V} X_v$. Since the X_v are (trivially) compact, X is compact.

A point of X can be identified with a map $f\colon V \to [k]$. For every edge $e = \{u, v\} \in E$, let $F_e \subseteq X$ consist of those maps $f\colon V \to [k]$ for which $f(u) \neq f(v)$. We want to prove that $\bigcap_{e \in E} F_e \neq \emptyset$.

What we know is that whenever $E_0 \subseteq E$ is a finite set of edges, we have $\bigcap_{e \in E_0} F_e \neq \emptyset$. This is because the finite graph consisting of the edges of E_0 and their vertices is assumed to be k-colorable.

By the definition of the product topology, it is easy to see that every F_e is closed. So it suffices to verify the following claim: *If $\mathcal{F}$ is a collection of closed sets in a compact space X such that every finite subcollection has a nonempty intersection, then $\mathcal{F}$ has a nonempty intersection.* But this is a reformulation of the definition of compactness—just consider $\mathcal{U} := \{X \setminus F : F \in \mathcal{F}\}$. □

Compact subsets of $\mathbb{R}^n$. Now we can easily establish the following well-known characterization.

Theorem 3.9 (Heine–Borel). *A subset $A \subseteq \mathbb{R}^n$, where $\mathbb{R}^n$ is equipped with the standard topology, is compact if and only if it is both closed and bounded.*

Proof. First we assume A compact. Then A is closed by use of Lemma 3.3(ii), and boundedness follows by considering the open cover by balls $B(0, n)$, $n = 1, 2, \ldots$.

For the other direction, it suffices to prove that the cube $[-m, m]^n$ is compact for every m, n, since then the case of a general A follows by Lemma 3.3(i).

The crucial part is in proving the interval $[0, 1]$ compact; the rest follows by re-scaling and by Tychonoff's theorem. The compactness of closed intervals is built deeply in the construction of the reals, and it is more or less a rephrasing of the fact that every subset of $\mathbb{R}$ has a supremum.

So let $\mathcal{U}$ be an open cover of $[0,1]$, and let s be the supremum of those $a \leq 1$ for which $[0,a]$ can be covered by finitely many members of $\mathcal{U}$.

Clearly $s > 0$. If $0 < s < 1$, then there is $\varepsilon > 0$ such that $[s-\varepsilon, s+\varepsilon]$ is covered by some $U \in \mathcal{U}$. Together with the assumed finite cover of $[0, s-\varepsilon]$, this U forms a finite cover of $[0, s+\varepsilon]$—a contradiction. □

Exercise 3.10. The previous result shows, in particular, that the Euclidean unit ball in $\mathbb{R}^n$ is compact.

(a) Consider the (infinite-dimensional Hilbert) space ℓ_2 consisting of all infinite sequences $x = (x_1, x_2, \ldots)$ of real numbers such that $\|x\| := \left(\sum_{i=1}^{\infty} x_i^2\right)^{1/2}$ is finite. Regard it as a topological space with topology induced by $\|.\|$, i.e., by the metric given by $d(x,y) = \|x-y\|$. Show that the unit ball $\{x \in \ell_2 : \|x\| \leq 1\}$ is not compact.

(b) Explain where the proof above, showing that B^n is compact, fails for the unit ball in ℓ_2. ✠

Paracompactness. There are many variations on compactness, most of them weaker than compactness, and none as significant. We mention just one notion, paracompactness, which often occurs among assumptions in other fields of mathematics.

We do not give the standard definition but an equivalent property which is most often used in applications. So let us assume that X is a Hausdorff space; then X is **paracompact** if every open cover $\mathcal{U}$ of X admits a partition of unity subordinated to $\mathcal{U}$. Here a *partition of unity subordinated to* $\mathcal{U}$ is a collection, finite or infinite, $(f_i)_{i \in I}$ of continuous functions $f_i \colon X \to [0,1]$ such that, first, for every $x \in X$, the sum $\sum_{i \in I} f(x)$ has only finitely many nonzero terms and equals 1, and second, for every $i \in I$ there is $U \in \mathcal{U}$ such that f_i is zero everywhere outside U.

Partitions of unity are a useful technical tool for gluing "locally defined" objects on X into a global object. Paracompactness is a relatively weak property: in particular, every compact space is paracompact, and all metric spaces are paracompact (which is a hard

result). A non-paracompact example is the long ray introduced in (F) above.

4. Homotopy and Homotopy Equivalence

So far we have considered two topological spaces equivalent (the same) if they are homeomorphic. But finding out whether two given spaces are homeomorphic is a very ambitious and generally hopeless task, since it is known that the algorithmic problem, given two spaces X and Y, decide whether $X \cong Y$, is *algorithmically unsolvable* (undecidable) in general. At the same time, homeomorphism can be decided in many specific settings, and topology is full of remarkable results of this kind. For example, later we will see that $\mathbb{R}^m \not\cong \mathbb{R}^n$ for $m \neq n$, which is well known but quite nontrivial.

Even stronger undecidability claims hold; for example, it is undecidable whether a given space X is homeomorphic to the 5-dimensional sphere S^5, a very simple-looking space.

An attentive reader might wonder how a topological space, a highly infinite object in general, is given to an algorithm that can accept only finite inputs. This question will be discussed later, but for the moment, one may think of the input X to the question of homeomorphism with S^5 as a space living in some $\mathbb{R}^n$ and built of finitely many 5-dimensional Lego cubes, for example.

Algebraic topology, a branch which we are now slowly entering, considers topological spaces with a coarser equivalence, called *homotopy equivalence*. For example, as we will see, all of the spaces $\mathbb{R}^n$, $n = 1, 2, \ldots$, are homotopy equivalent, and actually homotopy equivalent to a one-point space.

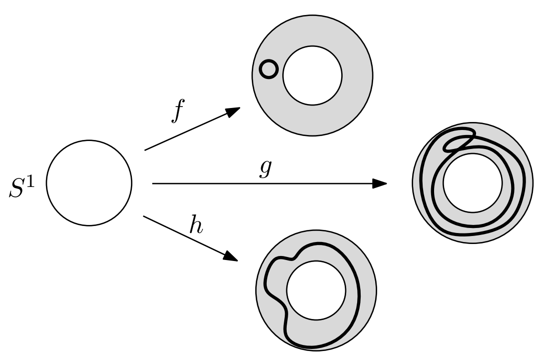
While deciding homotopy equivalence is still undecidable in general, in many concrete cases it is considerably easier than for homeomorphism. The reason is that there are many wonderful tools (the reader may have heard keywords like fundamental group, homotopy groups, homology and cohomology groups, etc.) that cannot distinguish between two homotopy equivalent spaces, but they can often prove homotopy non-equivalence.

Homotopy of maps. Homotopy equivalence is a somewhat sophisticated concept, which needs some time to be digested. We begin with an analogous but simpler notion for maps.

Definition 4.1. *Two (continuous) maps $f, g\colon X \to Y$ between the same spaces are called* **homotopic***, written $f \sim g$, if there exists a continuous map $H\colon X \times [0,1] \to Y$, a* homotopy *between f and g, satisfying $H(x,0) = f(x)$ and $H(x,1) = g(x)$ for all $x \in X$.*

Intuitively, f and g are homotopic if f can be continuously deformed into g (or g into f, which is equivalent). The homotopy H specifies such a deformation: we can think of the second coordinate t as time, and for every point $x \in X$, the map $h_x(t) = H(x,t)$ specifies the trajectory of the image of x during the deformation: it starts in $f(x)$ at time $t = 0$, moves continuously, and reaches $g(x)$ at time $t = 1$. The continuity of H implies that this trajectory is continuous for every x, and also that close points must have close trajectories. (Note that the continuity of H is in general stronger than just assuming that H is continuous in each of the coordinates; i.e., that $h_x(t)$ is a continuous function of t for every x, and that it is also a continuous function of x for every t.)

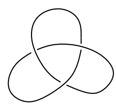
The next picture shows three maps of S^1 into the annulus (a part of the plane with a hole).



We have $f \sim g$ (imagine an appropriate deformation). But h is not homotopic to either of f, g—this is quite intuitive, since h goes once around the hole, while f and g do not go around, in a suitably

defined sense, but proving it rigorously is nontrivial, and we will leave it without proof for now.

Exercise 4.2. (a) Is the map $f\colon S^1 \to \mathbb{R}^3$ that maps S^1 to a geometric circle homotopic to a map $g\colon S^1 \to \mathbb{R}^3$ sending the circle to a knot, such as the trefoil? Answer before reading further!



(b) Let X be a space. Prove that every two maps $X \to B^n$ are homotopic.

(c) Prove that every two maps $B^n \to X$ are homotopic, provided X is path-connected. ✠

It is not difficult to show that being homotopic is an equivalence relation (writing down the proof of transitivity may take some work, but the idea is absolutely straightforward). We write $[X, Y]$ for the set of all homotopy classes of continuous maps $X \to Y$.

While there are usually uncountably many maps $X \to Y$, $[X, Y]$ is countable for spaces normally encountered in applications, sometimes even finite, and in many cases of interest it is well understood.

As a simple example we mention, again without proof, that the homotopy classes of maps of S^1 into the annulus are in a bijective correspondence with $\mathbb{Z}$, where each map is assigned the number of times the image winds around the hole, in positive (counterclockwise) or negative (clockwise) direction.

A map homotopic to a constant map $X \to Y$ (i.e., map all of X to a single point) is called—slightly illogically—**nullhomotopic**.

Homotopy equivalence. Now we come to spaces. The usual definition of homotopy equivalence is not very intuitive but good to work with.

Definition 4.3. *Two spaces X and Y are* **homotopy equivalent**, *written $X \simeq Y$, if there are continuous maps $f\colon X \to Y$ and $g\colon Y \to X$ such that the composition $fg\colon Y \to Y$ is homotopic to the identity map* id_Y *and gf is homotopic to* id_X.

The map g in the definition is called a *homotopy inverse* to f (and vice versa).

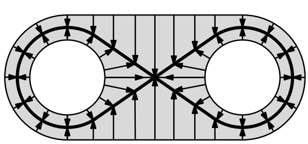
Similar to homotopy of maps, it is a simple exercise to show that homotopy equivalence is transitive. An equivalence class of spaces under homotopy equivalence is called a **homotopy type**.

Exercise 4.4. (a) Show that the dumbbell ⚬-⚬ and the letter θ are homotopy equivalent.

(b) (This is a very basic fact.) Check that $\mathbb{R}^n \setminus \{0\} \simeq S^{n-1}$. ✠

A way of visualizing homotopy equivalence uses the notion of *deformation retract*. Let X be a space and Y a subspace of X (this is important). A *deformation retraction* of X onto Y is a continuous map $R\colon X \times [0,1] \to X$ such that $R(.,0)$ is the identity map id_X, $R(t,y) = y$ for all $y \in Y$ and all $t \in [0,1]$ (Y remains pointwise fixed), and $R(x,1) \in Y$ for all $x \in X$. We say that Y is a **deformation retract** of X if there is a deformation retraction as above.

The deformation retraction R describes a continuous motion of points of X within X such that every point ends up in Y and Y remains fixed all the time. Here is an example, with X a thick figure 8 and Y a thin one:

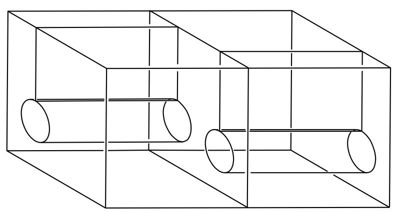


It turns out that two spaces X, Y are homotopy equivalent if and only if there exists a space Z such that both X and Y are deformation retracts of Z; see [**Hat01**, Corol. 0.21]. The direction which helps us with visualization, i.e., being deformation retracts of the same space implies homotopy equivalence, is exercise-level, and the other, with a right idea, is simple as well.

Exercise 4.5. Take an S^2 in $\mathbb{R}^3$ and connect the north and south poles by a segment, obtaining a space X. Take another copy of S^2 and attach a circle S^1 to the north pole by a single point, which yields Y. Show that $X \simeq Y$ (you may use deformation retracts). ✠

A space that is homotopy equivalent to a single point is called **contractible**.

Some spaces are "obviously" contractible, such as the ball B^n, but for others, contractibility is not easy to visualize. An example is **Bing's house**, one of the puzzling and beautiful objects of topology:



Bing's house is a hollow box with a wall inside separating it into two rooms, left and right. Each room has its own entrance, but by the architect's caprice, the entrance to the right room goes through a tunnel inside the left room (but is not accessible from the left room), and vice versa. Each of the tunnels is also attached to the ceiling by a vertical wall, which assures contractibility.

To check contractibility, one can visualize a deformation retraction of a solid cube onto Bing's house. If the cube is made of clay, one can push in a hole from the left and hollow out the right room through the hole, and similarly for the left room.

5. The Borsuk–Ulam Theorem

Here we interrupt our gradual introduction of basic topological notions and ideas, presenting the Borsuk–Ulam theorem, which is arguably one of the most useful tools topology has to offer to non-topologists. (Another theorem of comparable fame and usefulness is Brouwer's, which we will treat later.)

We begin by stating three versions, easily seen to be equivalent. The following notion will be useful: Let $X \subseteq \mathbb{R}^m$ and $Y \subseteq \mathbb{R}^n$ be antipodally symmetric sets; that is, $x \in X$ implies $-x \in X$. We call a continuous map $f\colon X \to Y$ an **antipodal map** if $f(-x) = -f(x)$ for all $x \in X$ (so an antipodal map is automatically assumed continuous).

Theorem 5.1 (Borsuk–Ulam)**.**

(i) *For every continuous map* $f\colon S^n \to \mathbb{R}^n$ *there is a point* $x \in S^n$ *with* $f(x) = f(-x)$.

(ii) *Every antipodal map* $g\colon S^n \to \mathbb{R}^n$ *maps some point* $x \in S^n$ *to* 0*, the origin in* $\mathbb{R}^n$.

(iii) *There is no antipodal map* $S^n \to S^{n-1}$.

Exercise 5.2. Prove the equivalence (i) $\Leftrightarrow$ (ii) $\Leftrightarrow$ (iii). ✠

Exercise 5.3. (Harder) Derive the following from Theorem 5.1: An antipodal map $S^n \to S^n$ cannot be nullhomotopic. ✠

The Borsuk–Ulam theorem comes from the 1930s and many different proofs are known. Unfortunately, conceptual proofs providing deeper insight require topological machinery beyond our scope, and the more elementary proofs we are aware of are often nice and clever, but one needs to spend considerable time with inessential technicalities. So we refer to the literature for a proof (e.g., [**Mat03**] or references therein), and instead we derive yet another, equivalent but different-looking, version.

Theorem 5.4 (Lyusternik–Schnirel'man)**.** *Let* $A_1, \ldots, A_{n+1} \subseteq S^n$ *be* $n+1$ *sets that together cover* S^n*, and let us assume that, for each* i*,* A_i *is either open or closed. Then some* A_i *contains a pair of antipodal points,* x *and* $-x$.

This theorem is traditionally presented either with all A_i closed or all A_i open, but allowing for a mixture can be useful, as we will see.

Exercise 5.5. (a) Construct a covering of S^n with $n+2$ closed sets, none containing an antipodal pair.

(b) Cover S^n with two sets, neither containing an antipodal pair. ✠

Proof of Lyusternik–Schnirel'man from Borsuk–Ulam. First we assume that all the A_i are closed, and we define a continuous map $f\colon S^n \to \mathbb{R}^n$ by $f(x)_i = \operatorname{dist}(x, A_i)$, the Euclidean distance of x from A_i. By the Borsuk–Ulam theorem there is $x \in S^n$ with $f(x) = f(-x)$. If $f(x)_i = 0$ for some i, then $x \in A_i$ (here we use the closedness), as well as $-x \in A_i$, and we are done. If, on the other hand, $f(x)_i > 0$ for all i, then x and $-x$ do not belong to any of $A_1, \ldots, A_n$, and so they both lie in A_{n+1}, the set which was seemingly neglected in the definition of f.

Next, let the A_i be all open. It suffices to show that there are closed $F_1 \subset A_1, \ldots, F_{n+1} \subset A_{n+1}$ that together still cover S^n, since then we can use the version with the A_i closed.

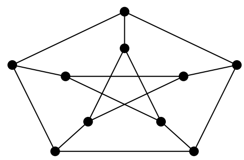
The proof of the last claim is a typical application of compactness. For every $x \in S^n$ we choose $i = i(x)$ such that $x \in A_i$, and an open neighborhood U_x of x whose closure is contained in $A_{i(x)}$. The U_x form an open cover of S^n, so we can choose a finite subcover, say $U_{x_1}, \ldots, U_{x_m}$. Then we set $F_i := \bigcup_{j:i(x_j)=i} \operatorname{cl} U_{x_j}$.

Finally, let $A_1, \ldots, A_k$ be open and $A_{k+1}, \ldots, A_{n+1}$ closed. We proceed by contradiction, supposing that no A_i contains an antipodal pair. Then, for each $i \geq k+1$, A_i has some positive distance $\varepsilon_i > 0$ from $-A_i$, and we let A'_i be the open $(\varepsilon_i/3)$-neighborhood of A_i. We still have $A'_i \cap (-A'_i) = \emptyset$, and hence the open sets $A_1, \ldots, A_k, A'_{k+1}, \ldots, A'_{m+1}$ contradict the version of the theorem for open sets proved above. □

Exercise 5.6. Derive the Borsuk–Ulam theorem from the Lyusternik–Schnirel'man theorem. Hint: use Exercise 5.5(a). ✠

Kneser graphs. For integers n and k, the **Kneser graph** $\mathrm{KG}_{n,k}$ has all k-element subsets of some fixed n-element set X as vertices. Two such subsets F_1, F_2 are connected by an edge in $\mathrm{KG}_{n,k}$ if they are disjoint.

A Kneser graph is typically quite large; it has $\binom{n}{k}$ vertices. As a small example, we note that $KG_{5,2}$ is isomorphic to the famous *Petersen graph*:



There are several reasons why Kneser graphs constitute an extremely interesting class of graph-theoretic examples. Perhaps the most remarkable property is that they have a significantly large chromatic number, but none of the non-topological methods for bounding the chromatic number from below can provide a reasonable lower bound.

We have already mentioned k-colorable graphs in connection with Proposition 3.8; here we just add that the **chromatic number** $\chi(G)$ of a graph G is the smallest k such that G is k-colorable.

The following celebrated result was conjectured by Kneser and proved by Lovász:

Theorem 5.7 (Lovász–Kneser)**.** *For $n \geq 2k$, we have $\chi(\mathrm{KG}_{n,k}) \geq n - 2k + 2$.*

The chromatic number of $\mathrm{KG}_{n,k}$ actually equals $n-2k+2$; finding a coloring is an elementary but nice exercise.

The perhaps most common general lower bound for the chromatic number is

$$\chi(G) \geq \frac{|V(G)|}{\alpha(G)},$$

where $\alpha(G)$, the *independence number* of G, is the size of a maximum independent set in G. This lower bound has a simple reason, since an equivalent definition of a k-colorable graph is that the vertex set can be covered by k independent sets.

Now $\mathrm{KG}_{n,k}$ has quite large independent sets, of size $\binom{n-1}{k-1}$, corresponding to the collection of all k-element sets containing a given point of the ground set. Setting $n = 3k - 2$, for example, we see that $\chi(\mathrm{KG}_{3k-2,k}) = k$, while the $|V(G)|/\alpha(G)$ lower bound yields less than 3.

Even more strongly, $\mathrm{KG}_{3k-2,k}$ also has the *fractional chromatic number* less than 3, where the fractional chromatic number $\chi_f(G)$ can be compactly defined as the infimum of fractions $\frac{a}{b}$ such that $V(G)$ can be covered by a independent sets so that every vertex is covered at least b times. The fractional chromatic number is an important graph parameter, and examples with a large gap between χ_f and χ are *very* rare.

Many proofs of the Lovász–Kneser theorem are known, but all of them are topological, or at least strongly inspired by the topological proofs. We present a particularly short and neat one.

Proof of the Lovász–Kneser theorem. The Kneser graph $\mathrm{KG}_{n,k}$ needs an n-element ground set X; we choose X as an n-point set in $\mathbb{R}^{d+1}$ in general position, where $d = n - 2k + 1$, and where general position means that no $d+1$ points of X lie on a common hyperplane passing through the origin.

For contradiction, we suppose that there is a proper coloring of $\mathrm{KG}_{n,k}$ by at most $n-2k+1 = d$ colors. We fix one such proper coloring and we define sets $A_1, \ldots, A_d \subseteq S^d$: For a point $x \in S^d$, we have $x \in A_i$ if there is at least one k-tuple $F \subset X$ of color i contained in the open halfspace $H(x) := \{y \in \mathbb{R}^d : \langle x, y \rangle > 0\}$ (i.e., x is a unit normal of the boundary of $H(x)$ and points into $H(x)$). Finally, we put $A_{d+1} = S^d \setminus (A_1 \cup \cdots \cup A_d)$.

Clearly, A_1 through A_d are open sets, while A_{d+1} is closed. By our version of the Lyusternik–Schnirel'man theorem, there exist $i \in [d+1]$ and $x \in S^d$ such that $x, -x \in A_i$.

If $i \leq d$, we get two disjoint k-tuples colored by color i, one in the open halfspace $H(x)$ and one in the opposite open halfspace $H(-x)$. This means that the considered coloring is not a proper coloring of the Kneser graph.

If $i = d+1$, then $H(x)$ contains at most $k-1$ points of X, and so does $H(-x)$. Therefore, the common boundary hyperplane of $H(x)$ and $H(-x)$ contains at least $n-2k+2 = d+1$ points of X, and this contradicts the choice of X. □

6. Operations on Topological Spaces

We have seen the *product* of topological spaces as an operation creating new spaces from old ones. Here we introduce some more operations.

Quotient. Given a topological space X and a subset $A \subset X$, we can form a new space by "shrinking A to a point." Two spaces can be "glued together" to form another space. A space can be factored using a group acting on it. Here is a general definition capturing all of these cases.

Definition 6.1. *Let X be a topological space and let $\approx$ be an equivalence relation on the set X. The points of the* **quotient space** $X/\!\approx$ *are the classes of the equivalence $\approx$, and a set $U \subseteq X/\!\approx$ is open if $q^{-1}(U)$ is open in X, where $q\colon X \to X/\!\approx$ is the* quotient map *that maps each $x \in X$ to the equivalence class $[x]_\approx$ containing it.*

If A is a subspace of X, one writes X/A for the quotient space $X/\approx$, where the classes of $\approx$ are A and the singletons $\{x\}$ for all $x \in X \setminus A$. This formalizes the "shrinking of A to a single point" mentioned above.

More generally, if $(A_i)_{i\in I}$ is a collection of disjoint subspaces, the notation $X/(A_i)_{i\in I}$ is used, with the expected meaning (each A_i is shrunk to a point).

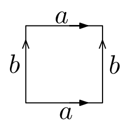
It is not hard to see, even rigorously, that $[0,1]/\{0,1\} \cong S^1$. Here are examples requiring more of mental gymnastics:

Exercise 6.2. Substantiate, at least on an intuitive level, the following homeomorphisms:

(a) $(S^n \times [0,1])/(S^n \times \{0\}) \cong B^{n+1}$.

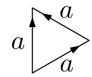
(b) $B^n/S^{n-1} \cong S^n$.

(c) $[0,1]^2/\!\approx \cong S^1 \times S^1$, where $\approx$ is given by the following identification of the sides of the square:



The picture means that each point of an arrow labeled a is to be identified with the corresponding point of the other a-arrow, and similarly for the b-arrows (so, in particular, all four corners are glued together). This is a well-known construction of the **torus**. ✠

The following identification of the sides of a triangle leads to a mind-boggling space called the **dunce hat**, with properties similar to those of Bing's house. The dunce hat can be made in $\mathbb{R}^3$, even from cloth, for example, but it is quite hard to picture mentally.

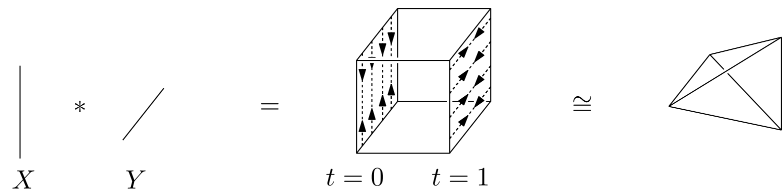


We should warn that if a quotient space is made in an irresponsible manner, we can obtain a badly-behaved topology even if we start with a nice space. For example, the quotient $\mathbb{R}^2/B^2$ can be shown to be homeomorphic to $\mathbb{R}^2$, but $\mathbb{R}^2/(\operatorname{int} B^2)$ is not even Hausdorff. Under normal circumstances, only *closed* subspaces should be shrunk to a point, but even that does not always guarantee good behavior.

If A is a closed subspace of X that is *contractible*, examples suggest that X/A should be homotopy equivalent to X (why not homeomorphic?). This, unfortunately, is *not* true in general, but it works for cases one is likely to encounter. Technically, an assumption guaranteeing that $X/A \simeq X$ for contractible A is called the *homotopy extension property* of the pair (X, A). We will not define it here; it suffices to say, with a forward reference to the next section, that if X is a simplicial or CW complex and A is a contractible subcomplex, then $X/A \simeq X$ holds.

Join. While various products and quotients are encountered in many mathematical structures, joins appear more specific to topology (joins in lattices or in database theory are similar to joins in topology only by name). The join $X * Y$ of spaces X and Y is obtained by taking the Cartesian product $X \times Y$, "fattening" it by another product with $[0, 1]$, and finally, collapsing the initial and final slices $X \times Y \times \{0\}$ and $X \times Y \times \{1\}$: in the former, each copy $X \times \{y\} \times \{0\}$ of X is collapsed to a point, while in the latter, the copies $\{x\} \times Y \times \{1\}$ of Y are collapsed. After these collapses, $X \times Y \times \{0\}$

becomes homeomorphic to Y, and $X \times Y \times \{1\}$ to X. Here is an illustration with X and Y segments:



The formal definition goes as follows.

Definition 6.3. *The* **join** $X * Y$ *of spaces* X *and* Y *is the quotient space* $(X \times Y \times [0,1])/{\approx}$, *where* $\approx$ *is given by* $(x, y, 0) \approx (x', y, 0)$ *for all* $x, x' \in X$ *and all* $y \in Y$ *("for* $t = 0$, x *does not matter") and* $(x, y, 1) \approx (x, y', 1)$ *for all* $x \in X$ *and all* $y, y' \in Y$ *("for* $t = 1$, y *does not matter").*

We observe that $X * Y$ contains the product $X \times Y$, e.g., as the "middle slice" $X \times Y \times \{\frac{1}{2}\}$. The join may look more complicated than the product, but in many respects it is better behaved; some of the advantages will be mentioned later.

There is a nice geometric interpretation of the join. Namely, suppose that X is represented as a bounded subspace of some $\mathbb{R}^m$, and Y of some $\mathbb{R}^n$. We then further insert $\mathbb{R}^m$ and $\mathbb{R}^n$ into $\mathbb{R}^{m+n+1}$ as skew affine subspaces, concretely $\{x \in \mathbb{R}^{m+n+1} : x_{n+1} = \cdots = x_{n+m+1} = 0\}$ and $\{y \in \mathbb{R}^{m+n+1} : x_1 = \cdots = x_n = 0, x_{n+1} = 1\}$ (so for $m = n = 1$ we have two skew lines in $\mathbb{R}^3$). With this placement of X and Y in $\mathbb{R}^{m+n+1}$ it can be verified that $X * Y$ is homeomorphic to the subspace $\bigcup_{x \in X, y \in Y} xy$ of $\mathbb{R}^{m+n+1}$, where xy is the segment connecting x and y. The point of placing X and Y into skew affine subspaces is to guarantee that two segments xy and $x'y'$, $x, x' \in X$, $y, y' \in Y$ never intersect, except possibly at one of the endpoints.

The join is commutative up to homeomorphism, but unfortunately not associative in general (although some of the literature claims so). For our purposes, though, it is amply sufficient that it is associative (up to homeomorphism of course) on the class of all compact Hausdorff spaces.

Cone and suspension. These are two popular special cases of the join. The **cone** of a space X is $\mathrm{C}X := X * \{p\}$, the join with a one-point space. Geometrically, the cone is the union of all segments connecting the points of X to a new point. We can also write $\mathrm{C}X$ as another quotient space, simpler than the one for a general join: $(X\times[0,1])/(X\times\{1\})$.

One of the simple ways of proving contractibility of a space Y is to show that Y is the cone of another space.

The join with a two-point space, $X * S^0$, is called the **suspension** of X and denoted by $\mathrm{S}X$. It can be interpreted as erecting a double cone over X. (Readers who find S^0 as two-point space puzzling may want to think it over—S^0 is used quite frequently.)

Exercise 6.4. (a) Show that $\mathrm{S}S^n \cong S^{n+1}$.

(b) Prove that $S^k * S^\ell \cong S^{k+\ell+1}$. Hint: use (a) and associativity of the join. ✠

While the cone operation makes every space homotopically trivial, i.e., contractible, the suspension more or less preserves the topological complexity, only pushing it one dimension higher. Very roughly speaking, it converts "k-dimensional holes" in X into "$(k+1)$-dimensional holes" in $\mathrm{S}X$.

6.1. Note on Categorical Definitions. The topology of the quotient $X/{\approx}$ can also be defined as the finest one for which the quotient map $q\colon X \to X/{\approx}$ is continuous. Here a topology $\mathcal{O}'$ is finer than $\mathcal{O}$ if $\mathcal{O} \subseteq \mathcal{O}'$. In the definition earlier, we described explicitly what the open sets are, but the formulation just given is equivalent.

The definition of the product topology on the Cartesian product $X := \prod_{i\in I} X_i$ in Section 3 can be rephrased similarly using the *projection maps* $p_i\colon X \to X_i$, where p_i maps an $|I|$-tuple $(x_i)_{i\in I} \in X$ to its i-th component x_i. Namely, *the product topology is the coarsest topology on X that makes all of the p_i continuous* (a topology $\mathcal{O}$ is coarser than $\mathcal{O}'$ if $\mathcal{O} \subseteq \mathcal{O}'$).

This is not only equivalent to the definition above Theorem 3.6 in Section 3, but it also explains one possibly ad-hoc looking aspect of

that definition, namely, why we admit only finitely many nontrivial factors in the open rectangles.

Exercise 6.5. Check the equivalence of both of the definitions of the product topology. ✠

Disjoint union. There is another, rather simple operation, which can be defined in a similar way. Namely, given a collection, finite or infinite, $(X_i)_{i\in I}$ of topological spaces, their **disjoint union** (or sometimes *disjoint sum*) $\coprod_{i\in I} X_i$ corresponds to the intuitive notion of putting disjoint copies of the X_i "side by side."

The ground set of $\coprod_{i\in I} X_i$ is the disjoint union of the sets X_i. Concretely, we may take $\bigcup_{i\in I} X_i \times \{i\}$, so that the elements of X_i are marked with i. This time we have the *inclusion maps* $\iota_i\colon X_i \to \coprod_{i\in I} X_i$, and the topology of the disjoint union is the finest one making all the ι_i continuous. Of course, it is not hard to describe the open sets explicitly as well: a set in $\coprod_{i\in I} X_i$ is open exactly if its intersection with each X_i is open.

The categorical approach. Here "categorical" is not related to Immanuel Kant but rather to the mathematical field of *category theory*, which studies general abstract structures in all of mathematics.

Why do we feel obliged to say something about categories in an introductory text on topology? First, category theory was invented by algebraic topologists, it has greatly helped cleaning up some unmanageably complicated, and thus potentially wrong, proofs in topology, facilitated much progress in the field, and it is heavily used in topology both as a language and as a tool.

Second, even if one does not intend to learn much about category theory, there are several basic principles definitely worth knowing about. In almost any field of mathematics or computer science, even a little bit of category-theory thinking can prevent one from re-inventing the wheel, or from riding on octagonal wheels where round ones are available.

Objects and morphisms. One of the starting points of category theory is that *maps* between mathematical objects deserve at least equal status as the objects. Moreover, knowing all maps into an

object and from it often gives enough information about the object, so that we need not consider the object's internal structure at all.

For example, in the category Top of topological spaces, we take all topological spaces as objects. We do not consider just any old maps between spaces, but the "right" structural maps, namely, all *continuous maps*.

In category theory, the maps of the "right kind" for a given type of objects are called **morphisms**. When studying some type of mathematical objects, what the morphisms are is not God-given, but to be user-defined. For many standard cases the morphisms are clear. For the category Set of sets they are arbitrary maps, for the category Grp of groups they are group homomorphisms, and for the category Gra of (simple, undirected) graphs they are graph homomorphisms.

Exercise 6.6. Recall as many mathematical structures as you can. In each case, try to find a reasonable notion of morphism between them. ✠

The next conceptual step in creating the category Top of topological spaces is to forget what are the ground set and open sets of each space, and where individual points are sent by the various maps. What is left? Well, a (tremendously infinite) *directed multigraph*. The spaces are the vertices, and each morphism (continuous map) $f\colon X \to Y$ gives rise to one arrow from X to Y. Importantly, information about *composition of morphisms* is also retained: given two arrows $f\colon X \to Y$ and $g\colon Y \to Z$, we know which of the arrows $X \to Z$ corresponds to the composition gf.

In general, a category is just that, a directed multigraph with an associative composition rule (or, if you prefer an algebraic language, a partial monoid). In more detail, a **category** C consists of the following data:

- A class[3] $\mathrm{Ob}(\mathsf{C})$ of objects.

[3] The objects in a category cannot be assumed to form a set set because of Russell's paradox. For example, if every set is an object of C, we cannot form the set of all sets, as Russell tells us. This is why the word *class* is used. Informally, a class is "like a set but possibly bigger"; for a mathematical foundation for working with classes see, e.g., [**AHS06**]. Categories whose class of objects is a set are called *small*.

- For every two objects $X, Y \in \mathrm{Ob}(\mathsf{C})$, a class $\mathrm{Hom}(X,Y)$ of morphisms from X to Y (with $\mathrm{Hom}(X,Y) \cap \mathrm{Hom}(U,V) = \emptyset$ whenever $(X,Y) \neq (U,V)$).
- For every $X \in \mathrm{Ob}(\mathsf{C})$, a unique *identity morphism* $\mathrm{id}_X \in \mathrm{Hom}(X,X)$.
- A *composition law* assigning to every $f \in \mathrm{Hom}(X,Y)$ and $g \in \mathrm{Hom}(Y,Z)$ an $h \in \mathrm{Hom}(X,Z)$, written as $h = gf$.

The composition is required to be associative, $f(gh) = (fg)h$, and satisfies $f\,\mathrm{id}_X = \mathrm{id}_Y\, f = f$ for every $f \in \mathrm{Hom}(X,Y)$.

Surprisingly many properties and constructions can be expressed solely in terms of objects and morphisms. Take the concepts of injectivity, surjectivity, and isomorphism. In category theory, the counterparts are:

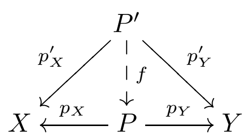
- A **monomorphism**, which is a left-cancellable morphism $f\colon X \to Y$, in the sense that $fg_1 = fg_2$ implies $g_1 = g_2$ for any two morphisms into X.
- An **epimorphism** is a right-cancellable morphism $f\colon X \to Y$, with $g_1 f = g_2 f$ implying $g_1 = g_2$.
- An **isomorphism** is a morphism $f\colon X \to Y$ that has a two-sided inverse; i.e., $g\colon Y \to X$ with $fg = \mathrm{id}_Y$ and $gf = \mathrm{id}_X$. An isomorphism is both a monomorphism and an epimorphism, but these conditions are not sufficient in general.

Exercise 6.7. (a) Check that in the category Set, monomorphisms and epimorphisms correspond to injective and surjective maps, respectively.

(b) Consider the category Haus of all *Hausdorff* topological spaces with continuous maps as morphisms. Let us consider the rationals $\mathbb{Q}$ as a subspace of $\mathbb{R}$ with the standard topology, and let $f\colon \mathbb{Q} \to \mathbb{R}$ be the standard inclusion. Is f an epimorphism? Can you characterize what epimorphisms are in this category? ✠

Products revisited. Products, for example, have a general categorical definition. Given objects X and Y in a category C, this definition identifies the product of X and Y, if one exists, up to isomorphism.

Namely, the **product** $X \times Y$ in C is an object P *plus* morphisms $p_X \colon P \to X$ and $p_Y \colon P \to Y$ with the following *universal property*: whenever P' is an object and $p'_X \colon P' \to P$ and $p'_Y \colon P' \to Y$ are morphisms, there is a unique morphism $f \colon P' \to P$ with $p'_X = p_X f$ and $p'_Y = p_Y f$. Or, expressed in a way category theorists and topologists prefer, there is a unique f making the following diagram commutative:



It is easy to see that such a P, if it exists, is unique up to isomorphism. The definition for the product of arbitrarily many objects is entirely analogous. As we have already indicated, not every category has products, but many do.

This definition may very well look nonintuitive and difficult to work with, and certainly it takes time and training to get used to that kind of reasoning. For specific categories, it may take some work to figure out what the product "looks like." On the other hand, the categorical approach maintains that once we know that a product exists, the defining property above is the only one we really need for working with it, and that we may never *need* to figure out the specific structure, especially if we are working in some less common category.

Exercise 6.8. (a) Check that the product of topological spaces satisfies the categorical definition.

(b) Take Gra, graphs with graph homomorphisms. Describe the categorical product (for two graphs) concretely, in terms of vertices and edges.

Limits. The product construction is a special case of **categorical limit**. That definition tells us what is the limit of a given (commutative) diagram in a given category C. Since we do not want to define diagrams in general, let us give just an example.

We consider three objects A, X, Y with morphisms $f\colon X \to A$ and $g\colon Y \to A$. The limit of the diagram

$$\begin{array}{ccc} & & X \\ & & \downarrow{\scriptstyle f} \\ Y & \xrightarrow{g} & A \end{array}$$

is an object T plus morphisms $p_X : T \to X$ and $p_Y : T \to Y$ making the following digram commutative

$$\begin{array}{ccc} T & \xrightarrow{p_X} & X \\ {\scriptstyle p_Y}\downarrow & \searrow^{p_A} & \downarrow{\scriptstyle f} \\ Y & \xrightarrow{g} & A \end{array}$$

and satisfying the universality property: whenever T' and p'_A, p'_X, p'_Y is another completion to a commutative diagram, there is a unique morphism $u\colon T' \to T$ such that $p'_A = p_A u$, $p'_X = p_X u$, and $p'_Y = p_Y u$.

For this particular diagram, the limit is called the *pullback*.

The same definition of a limit works for any commutative diagram in C; the morphisms p_X go from the limit object to every object in the diagram. The product is the special case of a limit where the diagram has just objects and no morphisms.

Exercise 6.9. Work out what the pullback looks like in Set. ✠

Opposite category and conotions. For every category C we can immediately form a new category C^{op} by reversing all arrows. This, of course, would be highly problematic for actual maps, since how should one invert a map that is not bijective, but it is no problem for a category theorist, who regards morphisms as abstract arrows.

For every categorical notion, we can form a "dual" notion by reversing all arrows. From product we get **coproduct**, which for topological spaces turns out to be just the disjoint union. (Here and in many other categories, the coproduct is rather dull, but for example, in the groups category Grp it is the *free product* of groups.) From limit we get *colimit*, etc., the prefix co- expressing the dual nature of the notion. (This terminology has some common sense exceptions, such as epimorphism instead of comonomorphism and *pushout* instead of

copullback. But physicists may have missed an opportunity here with their *bra* and *ket* terminology.)

Category theory has a number of general constructions and theorems, and many concrete constructions get simplified by observing that they are but special realizations of these general abstract results. In topological and other proofs, references to such general categorical considerations are often (proudly) prefixed by the phrase "by abstract nonsense it follows that...."

7. Simplicial Complexes and Relatives

7.1. Simplicial Complexes and Simplicial Maps. We have already touched upon the question: how can interesting topological spaces be described in a finite way? Simplicial complexes provide the simplest systematic way. Real topologists often frown on them and consider them old-fashioned as a theoretical tool and not economical enough compared to other tools. These are perfectly valid concerns, but for computer-science and combinatorial uses, simplicial complexes may often be the winners because of their combinatorial simplicity.

As a combinatorial object, a simplicial complex is simply a hereditary system of finite sets:

Definition 7.1. *A* **simplicial complex** *is a system K of finite subsets of a (possibly infinite) set V, with the property that if $F \in K$ and $F' \subset F$, then $F' \in K$ as well. The set V, called the* **vertex set** *of K and denoted by $V(K)$, is the union of all sets of K.*

In rare cases, it may be useful to also admit, unlike in the definition above, points of V that do not belong to any $F \in K$.

The definition implies, in particular, that $\emptyset \in K$ whenever $K \neq \emptyset$; in some of the literature, though, the empty set is not regarded as a member of K.

The sets in K are called the **simplices** of K. The vertex set is sometimes also called the *ground set*.

There is some formal ambiguity in using the term **vertex** of a simplicial complex: it may mean a point v of the vertex set V or a

singleton set $\{v\}$, which is a simplex of K. But in practice this does not lead to confusion.

A **subcomplex** of a simplicial complex K is a simplicial complex $L \subseteq K$. We say that L is an **induced subcomplex** of K if $L = \{F \in K : F \subseteq V(L)\}$, i.e., every simplex of K living on the vertex set of L also belongs to L.

The **dimension** of a simplicial complex K is

$$\dim K := \sup_{F \in K} (|F| - 1).$$

The "-1" in this definition is logical, of course, since, e.g., a three-point $F \in K$ will correspond to a geometric triangle, which is 2-dimensional, but it is an eternal source of confusion.

A useful example to keep in mind are 1-dimensional simplicial complexes, which can be regarded as simple graphs: the 0-dimensional simplices correspond to vertices and 1-dimensional ones to edges. Historically, the study of graphs has for some time been regarded as a part of topology.

Finite and infinite simplicial complexes. A simplicial complex is **finite** if it has a finite ground set. By definition, a simplicial complex can also be infinite, for a good reason: as we will see, finite simplicial complexes can describe only compact subspaces of some $\mathbb{R}^n$, which excludes spaces like $(0, 1)$ or $\mathbb{R}^n$ itself.

On the other hand, only finite simplicial complexes can naturally serve as inputs to algorithms, which was one of our main motivations for considering simplicial complexes. Moreover, for many purposes, including most of computer-science related applications, finite simplicial complexes suffice. Infinite simplicial complexes originally served as a theoretical tool for building algebraic topology, but in that role they have been replaced by more modern tools.

We will restrict ourselves to finite simplicial complexes, except for a couple of remarks.

Simplicial maps. By now the reader may be impatient to see what is the topological space described by a simplicial complex, but before explaining that, we will still want to say what are the appropriate

maps (*morphisms* in the categorical jargon newly introduced above) between simplicial complexes.

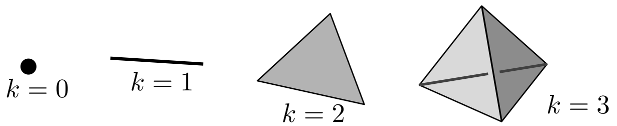
Definition 7.2. *A* **simplicial map** *of a simplicial complex K into a simplicial complex L is a map $s\colon V(K) \to V(L)$ of the vertex sets that maps simplices to simplices, i.e., $s(F) \in L$ for every $F \in K$. An* **isomorphism** *of simplicial complexes is a bijective simplicial map with simplicial inverse.*

Isomorphism, similar to many other mathematical structures, means that the simplicial complexes have identical structure and differ only by renaming vertices.

We note that simplicial maps for 1-dimensional simplicial complexes are not the same as graph homomorphisms, since unlike homomorphisms, they allow for edges to be collapsed to vertices. But isomorphism is the same notion for graphs and 1-dimensional simplicial complexes.

7.2. Geometric Realization and Polyhedra. Now we want to say what the topological space described by a (finite) simplicial complex K is.

First we recall that a (geometric) **simplex** is the convex hull of a set of affinely independent points[4] in some $\mathbb{R}^n$; simplices of dimension 0, 1, 2, 3 are points, segments, triangles, and tetrahedra, respectively.



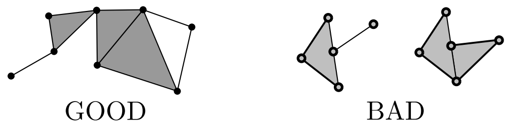
The *faces* of a simplex σ are the convex hulls of subsets of the vertex set. For example, a tetrahedron has 16 faces: itself, 4 triangles, 6 edges, 4 vertices, and the empty set. A k-dimensional simplex has $k + 1$ faces of dimension $k - 1$.

[4] Points $p_0, p_1, \ldots, p_k \in \mathbb{R}^n$ ($k + 1$ of them) are called **affinely independent** if the k vectors $p_1 - p_0, \ldots, p_k - p_0$ are linearly independent.

Definition 7.3. *A* **geometric simplicial complex** *is a collection* Δ *of geometric simplices of various dimensions satisfying the following two conditions:*

(i) *(Hereditary) If* $\sigma \in \Delta$ *and* σ' *is a face of* σ, *then* $\sigma' \in \Delta$.
(ii) *(Intersecting in faces) For every* $\sigma, \sigma' \in \Delta$, $\sigma \cap \sigma'$ *is a face of both* σ *and* σ'.

Somewhat informally, the simplices in a geometric simplicial complex may be glued only along common faces:



A geometric simplicial complex Δ defines a simplicial complex $K = K(\Delta)$ in the sense of Definition 7.1 in an obvious way: we set $V(K) = V(\Delta)$, the latter denoting the set of all vertices of the simplices in Δ, and the simplices of K are vertex sets of the simplices in Δ.

Now the geometric simplicial complex Δ is called a **geometric realization** of this K, and also of any simplicial complex K' isomorphic to K.

Proposition 7.4. *Every finite simplicial complex* K *has a geometric realization; if* $k = \dim K$ *then the realization can be taken in* $\mathbb{R}^{2k+1}$.

Sketch of proof. A geometric realization of K in some $\mathbb{R}^n$ is fully specified by the placement of the vertex set. Thus, we seek an (injective) map $\rho\colon V(K) \to \mathbb{R}^{2k+1}$.

The condition we need is that, for every two simplices $F, G \in K$, $\mathrm{conv}(\rho(F)) \cap \mathrm{conv}(\rho(G)) = \mathrm{conv}\,\rho(F \cap G)$, where conv(.) denotes the convex hull. A sufficient condition for this is that $\rho(F \cup G)$ be affinely independent, since then $\mathrm{conv}\,\rho(F \cup G)$ is a geometric simplex, both $\mathrm{conv}\,\rho(F)$ and $\mathrm{conv}\,\rho(G)$ are faces of it, and they intersect in the (possibly empty) face $\mathrm{conv}\,\rho(F \cap G)$ as they should.[5]

[5] Obvious as it may seem, this fact still needs a little proof, which we allow ourselves to omit. Here we are basically asserting that the set of all faces of a geometric simplex constitutes a geometric simplicial complex.

So it suffices to show that for every n there is an n-point set in $\mathbb{R}^{2k+1}$ in which every $2k+2$ points are affinely independent (because $2k+2$ is the maximum possible size of $F \cup G$). This we leave as an exercise for the readers not familiar with the trick. □

Exercise 7.5. Verify that every $d+1$ distinct points on the *moment curve* $\{(t, t^2, \ldots, t^d) : t \in \mathbb{R}\} \subset \mathbb{R}^d$ are affinely independent. Hint: a polynomial of degree at most d has at most d roots. ✠

Now, finally, we define the space associated with a simplicial complex.

Definition 7.6. *Let Δ be a geometric simplicial complex, and suppose that all simplices of Δ are contained in $\mathbb{R}^n$. The* **polyhedron** *of Δ is the topological subspace of $\mathbb{R}^n$ induced by the union of all simplices of Δ. A* **polyhedron** *of a finite simplicial complex K is the polyhedron of a geometric realization of K.*

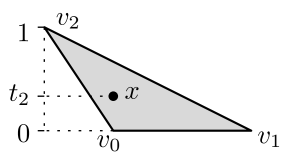
The polyhedron of K is not defined uniquely, but as we will soon see, all polyhedra of K are homeomorphic. The polyhedron of K is usually denoted by $|K|$, but often one writes K for the polyhedron as well, and one has to distinguish from the context whether the combinatorial object or the geometric one is meant.

Remark on infinite simplicial complexes. As we have mentioned above, defining the polyhedron of an infinite simplicial complex is somewhat more demanding. An immediate trouble is that all of the geometric simplices may not fit in the same $\mathbb{R}^n$, for example if the dimension is unbounded.

The solution uses quotient spaces. First, we assign a k-dimensional geometric simplex $\rho(F)$ to every k-dimensional $F \in K$, possibly each $\rho(F)$ in a different Euclidean space. Then we introduce a suitable equivalence relation $\approx$ on the disjoint union of these simplices, which amounts to identifying, for every $G \subset F$, the simplex $\rho(G)$ with the appropriate face of the simplex $\rho(F)$ (some care is needed in saying how exactly these identifications are performed; it is helpful to fix a linear ordering of the vertices of K first). Finally, $|K|$ is defined as the quotient of the disjoint union by $\approx$.

How simplicial maps yield continuous maps. Let K and L be simplicial complexes, and let $s\colon V(K) \to V(L)$ be a simplicial map. There is a canonical continuous map $|s|\colon |K| \to |L|$ of the polyhedra associated to s.

One often says that $|s|$ is a linear extension of s on the simplices of $|K|$ (although, strictly speaking, it is an affine extension). To define $|s|$ precisely, we need to recall that if σ is a geometric simplex with vertices $v_0, \ldots, v_k$, then every point $x \in \sigma$ can be uniquely written as $x = \sum_{i=0}^{k} t_i v_i$, where $t_0, \ldots, t_k \geq 0$ and $\sum_{i=0}^{k} t_i = 1$. Here $(t_0, \ldots, t_k)$ is called the *barycentric coordinates* of x; t_i is the height of x above the $(k-1)$-dimensional face of σ not containing v_i, scaled so that v_i has height 1:



So let us fix geometric realizations Δ and Δ' of K and L, respectively, and regard s as a map $V(\Delta) \to V(\Delta')$. For a point x in the polyhedron of Δ we choose a lowest-dimensional simplex σ containing x (such a σ is called the **support** of x in Δ and it is determined uniquely). We have $x = \sum_{i=0}^{k} t_i v_i$, where $v_0, \ldots, v_k$ are the vertices of σ, and we set

$$|s|(x) := \sum_{i=0}^{k} t_i s(v_i).$$

The sum is well defined because, by the definition of a simplicial map, $\{s(v_i) : i = 0, 1, \ldots, k\}$ is the vertex set of some simplex in Δ'.

Note that, since simplicial maps are allowed to map higher-dimensional simplices to lower-dimensional ones, the image of a k-dimensional simplex under $|s|$ may have any dimension $\ell \leq k$.

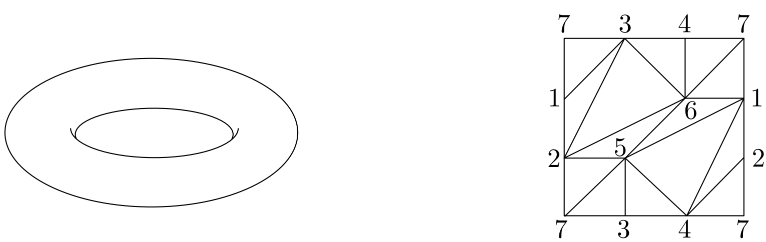
One needs to check that $|s|$ is continuous when we go from the interior of some simplex towards a point of a face, but this is straightforward.

It is also not hard to see that if s is injective, then so is $|s|$, and if s is an isomorphism, then $|s|$ is a homeomorphism. From this we immediately get that isomorphic simplicial complexes have homeomorphic polyhedra. In particular, the polyhedron of a simplicial complex is uniquely defined, up to a homeomorphism.

Triangulations. A simplicial complex K is called a **triangulation** of a space X if $X \cong |K|$. Naturally not all topological spaces possess a triangulation: some for reasons of local pathology, such as not being Hausdorff, but some others are non-triangulable in spite of being locally very nice. The perhaps most striking example is a 4-dimensional compact manifold (the *Freedman E8 manifold*; manifolds will be introduced later).

The simplest triangulation of the sphere S^{n-1} is the boundary of an n-dimensional simplex, with n simplices of dimension $n-1$ but $2^n - 1$ simplices in total. Combinatorially, denoting the vertices by $1, 2, \ldots, n$, the simplicial complex is $\{F \subseteq [n] : F \neq [n]\}$. Another, more symmetric triangulation will be mentioned soon.

It can be shown that every triangulation of the torus must have at least 7 vertices and at least 14 triangles, and here is one attaining these minimal numbers:



The triangulation is drawn as a square, but the sides of the square should be identified as in Exercise 6.2(c)—this is also indicated by the numbering of the vertices.

It may be worthwhile if the reader draws her own triangulation of a torus, trying to get a small number of triangles, and notes the pitfalls in such an enterprise.

The study of triangulations is a major and fast-growing area, but here we leave it aside, referring to [**DLRS10**] and [**Lut15**].

Simplicial joins. The join operation can also be done on the level of simplicial complexes in a straightforward way.

Let K, L be simplicial complexes, and first assume $V(K) \cap V(L) = \emptyset$. Then the simplicial join $K * L$ is $\{F \cup G : F \in K, G \in L\}$, on the vertex set $V(K) \cup V(L)$. If the vertex sets are not disjoint, we must first replace L, say, with an isomorphic simplicial complex whose vertex set is disjoint from $V(K)$.

It is not hard to show that $|K * L| \cong |K| * |L|$. The main step is in checking that the join of a geometric k-simplex and ℓ-simplex is a $(k + \ell + 1)$-simplex, which is easy using the interpretation of join with skew affine subspaces.

We saw (Exercise 6.4) that $S^n \cong (S^0)^{*(n+1)}$, the $(n+1)$-fold join of the 0-dimensional sphere, or two-point space. If we do this join simplicially, we obtain the following triangulation of S^n: the vertex set is $\{a_1, b_1, \ldots, a_{n+1}, b_{n+1}\}$, and a set of vertices forms a simplex exactly if it does not contain a pair $\{a_i, b_i\}$ for any i.

The geometric realization is the boundary of the so-called $(n+1)$-dimensional *crosspolytope*, a regular octahedron for $n = 2$ (just identify a_i with e_i, the i-th vector of the standard basis of $\mathbb{R}^{n+1}$, and b_i with $-e_i$). One often speaks of the *octahedral sphere*.

7.3. Combinatorial Examples. A great feature of simplicial complexes is that they give a way of assigning a topological space to all kinds of combinatorial objects: whenever we have a system of finite sets, we can close it under taking subsets, if it is not yet hereditary by itself, and we have a simplicial complex. Sometimes this topology connection is fruitful, sometimes not so much, but definitely there is something to study. We list several cases where this approach has been used with great success; many others can be found in the literature.

Clique (or flag) complexes of graphs. Given a (simple, undirected) graph G, we define a simplicial complex $C = C(G)$ on the vertex set $V(G)$ whose simplices are sets of vertices forming a clique (every two vertices connected by an edge). This C has several common names: **clique complex**, **flag complex**, *Whitney complex*, and

probably others. A similar complex $I(G)$ whose simplices are independent sets in G is the **independence complex** of G.

Clique complexes, besides constituting an interesting subclass of simplicial complexes, carry lot of information about the underlying graph in their topology. They feature, for example, in a proof by Meshulam [**Mes01**] of a generalization of a lovely theorem of Aharoni and Haxell [**AH00**], a Hall-type theorem for hypergraphs. The Aharoni–Haxell theorem was later used for proving a tantalizing combinatorial conjecture known as the *tripartite Ryser conjecture* [**Aha01**]; Kalai's blog [**Kal12**] provides a nice introduction. In computational geometry one finds a special case of the clique complex as the *Vietoris–Rips complex.*

Order complex. Let $(X, \preceq)$ be a partially ordered set. Its **order complex** lives on the vertex set X and the simplices correspond to *chains*, i.e., subsets of X linearly ordered by $\preceq$. (Equivalently, this is the clique complex of the comparability graph of $(X, \preceq)$.)

There is an extensive topological theory of order complexes; see, e.g., [**Bjö95, Wac07**]. For example, there is a famous fixed-point theorem for posets of Baclawski and Björner, which has a combinatorial statement but only topological proofs. Deep connections were found to questions in algebra and Lie theory.

Nerve. Let $\mathcal{F} = \{F_1, F_2, \ldots, F_n\}$ be a family of sets, so far arbitrary. The **nerve** $\mathcal{N}(\mathcal{F})$ of $\mathcal{F}$ is the simplicial complex with vertex set $[n]$, the set indices, and simplices corresponding to nonempty intersections:

$$\mathcal{N}(\mathcal{F}) = \Big\{ I \subseteq [n] : \bigcap_{i \in I} F_i \neq \emptyset \Big\}.$$

Here is a remarkable theorem.

Theorem 7.7 (Nerve theorem)**.** *Let $K_1, K_2, \ldots, K_n$ be subcomplexes of a finite simplicial complex K that together cover K (each simplex of K is in at least one K_i). Suppose that the intersection $\bigcap_{i \in J} |K_i|$ is empty or contractible for each nonempty $J \subseteq [n]$. Then the (polyhedron of the) nerve of $\{|K_1|, \ldots, |K_n|\}$ is homotopy equivalent to $|K|$.*

This often allows one to simplify a simplicial complex drastically while keeping the homotopy type.

There are many variations of the nerve theorem in the literature. For example, a useful geometric setting where a nerve theorem holds is when $F_1, \ldots, F_n$ are closed convex sets in a Euclidean space. This result, and various generalizations, are closely related to Helly-type theorems in geometry (see, e.g., [**CdVGG12**] for recent progress).

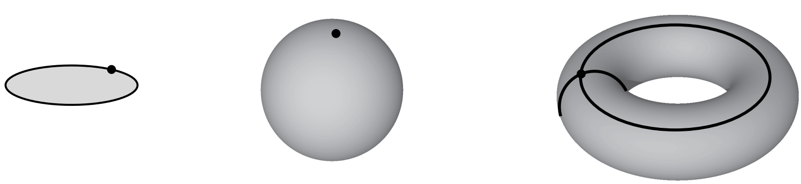
The usual proofs of nerve theorems use somewhat more machinery than we are going to develop here, so we refer, e.g., to Björner [**Bjö03**] for a relatively simple and elementary proof which also yields a powerful generalization.

7.4. Simplicial Sets and Cell Complexes. In geometric simplicial complexes, the simplices can be glued only in a somewhat rigid face-to-face manner. For example, we are not allowed to glue two triangles by two sides without actually making them identical, although geometrically, such a construction makes perfect sense. Other descriptions of spaces used in topology allow for more flexible gluing, and some of them use building blocks other than simplices.

Cell (or CW) complexes. This is perhaps the most popular way in topology, but it does not provide a finite description of a space, so we mention it only briefly.

The building blocks here are topological balls of various dimensions, called *cells*, which can be thought of as being completely "flexible" and which can be glued together in an almost arbitrary continuous fashion. Essentially, the only condition is that each n-dimensional cell has to be attached along its boundary to the $(n-1)$-*skeleton* of the space, i.e., to the part that has already been built, inductively, from lower-dimensional cells.

Here are some pictorial examples:

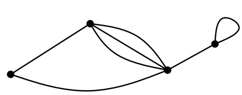
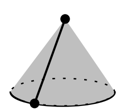
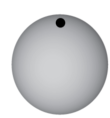


The left one is a B^2 made of a 0-cell (point), 1-cell, and 2-cell. The middle shows an S^2 made from one 0-cell and one 2-cell (whose boundary is shrunk to a point), and the right picture is the torus from a 0-cell, two 1-cells, and a single 2-cell.

Let us stress that the gluing of the n-dimensional cells to the $(n-1)$-skeleton is quite arbitrary. For example, gluing the boundary circle of a 2-cell, i.e., of a disk, to the 1-skeleton need not respect the vertices and edges of the 1-skeleton in any way. So if we like, we can identify two opposide half-circles of the boundary to a segment, and glue this segment to the middle of an edge. We refer to standard textbooks for a formal definition of a cell complex.

Simplicial sets. Although the name may suggest the opposite, simplicial sets are more complicated than simplicial complexes.

Intuitively, a simplicial set, similar to simplicial complexes, is made by gluing simplices of various dimensions together. The gluing is still face-to-face but much more permissive than for simplicial complexes. For example, one may have several 1-dimensional simplices connecting the same pair of vertices, a 1-simplex forming a loop, two edges of a 2-simplex identified to create a cone, or the boundary of a 2-simplex all contracted to a single vertex, forming an S^2.

The features mentioned so far are also shared by *Delta-complexes*, as used, e.g., in Hatcher [**Hat01**]. They ensure both the possibility of a finite combinatorial description (assuming finitely many simplices), and also considerable economy of description, at least in small cases: the torus can be described as a simplicial set with two triangles, as opposed to the minimum of 14 for simplicial complexes, or S^n needs only one 0-dimensional and one n-dimensional simplex, as opposed to at least 2^n simplices for a simplicial complex.

There is still another feature of simplicial sets, which looks peculiar at first sight: *degenerate simplices*. If σ is a simplex in a simplicial

set X, say two-dimensional with vertices v_0, v_1, v_2, then X also contains the degenerate simplex, denoted by $s_0\sigma$, which we can think of as simplex that geometrically coincides with σ but in which the vertex v_0 has been duplicated. The degenerate simplex $s_0\sigma$ is treated as 3-dimensional. Not only that, X also has to contain $s_1\sigma$, $s_2\sigma$, $s_0s_0\sigma$, and so on, degeneracies of σ of all possible orders.

We thus see that every nonempty simplicial set must have infinitely many simplices, which seems to ruin the purpose of finite description. Fear not: the degenerate simplices can be represented implicitly, since, as it turns out, every degenerate simplex can be specified by some nondegenerate starting simplex and a canonical sequence of the degeneracy ("vertex-duplicating" operators) s_i. So the degenerate simplices need not be stored explicitly. If a simplicial set has finitely many nondegenerate simplices, then it has a (quite efficient) finite encoding; this holds, in particular, when a finite simplicial complex is converted to a simplicial set.

After this misty introduction, what about a formal definition of a simplicial set? The preferred one in modern texts goes as follows: *A simplicial set is a contravariant functor* $\Delta \to \mathsf{Set}$, *where* Δ *denotes the category of nonempty finite linearly ordered sets, with (nonstrictly) monotone maps as morphisms.*

This kind of definition sounds somewhat cryptic, and such very compact and abstract definitions contribute to the difficult accessibility of modern algebraic topology. Even if one knew what a contravariant functor is (which will be mentioned here later on), it seems rather hard, without long training, to make any intuitive sense of this definition, and see how it may correspond to the things discussed informally above (of course, there may be exceptions among our readers).

Fortunately, there is a friendly pictorial treatment of simplicial sets by Friedman [**Fri12**], where one can also find a still rigorous but more descriptive reformulation of the above concise definition.

This section can be taken as an advertisement of simplicial sets, which, unlike simplicial complexes, seem to remain almost unknown in computer science. A great advantage compared to simplicial complexes is that the product of simplicial sets is again a simplicial set, and so is the quotient of a simplicial set by a simplicial subset.

One of remarkable uses of simplicial sets is theoretical. According to a theory worked out mainly by Kan, using sufficiently rich (necessarily infinite) simplicial sets, one can capture all homotopy classes of continuous maps between spaces by simplicial maps among the corresponding simplicial sets, similarly for homotopy classes of homotopies, etc. In this way, "continuous" homotopy theory can be imitated purely discretely in the category of simplicial sets with simplicial maps, and these ideas also allow one to do homotopy theory in, say, algebraic categories with no notion of continuity.

Simplicial sets do not appear outside topology with the spontaneity of simplicial complexes, yet they have also found impressive algorithmic uses. They constitute the main data structure in algorithms for computing homotopy groups or homotopy classes of continuous maps. We refer to [**RS12**] for nice lecture notes on this subject and to [**ČKM$^+$14**] for a sample of such algorithms.

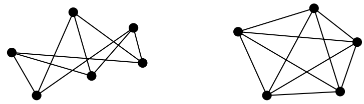
8. Non-embeddability

One of the first questions addressed in basic graph theory is, which graphs are planar? A planar graph is one that can be drawn in the plane without edge crossings. In topological terms, we have a 1-dimensional simplicial complex G and we ask whether its polyhedron $|G|$ can be embedded in $\mathbb{R}^2$.

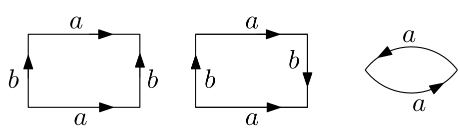
In general, if X and Y are topological spaces, an **embedding** of X in Y is a map $f : X \to Y$ that is a homeomorphism of X with the image $f(X)$. In other words, we are looking for a subspace of Y homeomorphic to X.

Let us remark that if X is compact and Y Hausdorff, as is the case for polyhedra of finite simplicial complexes and for finite graphs in particular, then we only need to look for an injective continuous map $X \to Y$; the inverse is continuous automatically.

Graph theory has a number of very satisfactory answers to the question above. One of them is Kuratowski's theorem, asserting that a graph is planar if and only if it does not contain a subdivision of $K_{3,3}$ or K_5, which are thus the basic nonplanar graphs:



The planarity question has obvious generalizations, where we consider higher-dimensional simplicial complexes K and we ask whether they embed in $\mathbb{R}^d$ for some given d. (Of course, one can be even much broader and go beyond Euclidean spaces as targets, etc.) One of the classical facts very often mentioned when discussing 2-dimensional surfaces is that while the torus, whose quotient-space construction is recalled below on the left, obviously embeds in $\mathbb{R}^3$, neither the *Klein bottle* (middle picture) nor the *projective plane* (right) do.



Various cases of the embeddability question has been an important topic in topology for many decades, and lots of very interesting partial results are known. However, it is clear that Kuratowski's theorem is a great positive exception—there cannot be any comparably conclusive results about embeddability of k-dimensional simplicial complexes in $\mathbb{R}^d$ for any $k \geq 2$. In particular, it is known that the algorithmic question, does a given finite 4-dimensional simplicial complex embed in $\mathbb{R}^5$, is *undecidable*, and so there is no hope at reasonable characterization theorems.

There is another point which should be stressed before we leave this general introduction. For planar graphs, a remarkable theorem of Fáry tells us that every planar graph has a straight-edge planar drawing, or in other words, that every 1-dimensional simplicial complex embeddable in $\mathbb{R}^2$ also has a geometric realization in $\mathbb{R}^2$, in which the simplices are straight.

For simplicial complexes of dimension 2 and higher, these two notions (embeddability in $\mathbb{R}^d$ and having a geometric realization in $\mathbb{R}^d$), are completely different in general. Here we will talk exclusively about (topological) embeddability.

In the rest of the section we want to demonstrate a single non-embeddability result, where we show an interesting general method, as well as some of the notions and tools introduced earlier in action.

Van Kampen–Flores complexes. We saw in Proposition 7.4 that every k-dimensional finite simplicial complex K embeds in $\mathbb{R}^{2k+1}$. Here we will prove a complementary and classical result.

Theorem 8.1 (van Kampen–Flores)**.** *For every integer $k \geq 1$, there are k-dimensional finite simplicial complexes that cannot be embedded in $\mathbb{R}^{2k}$.*

Thus, the bound of $2k+1$ of Proposition 7.4 cannot be improved in general. We can also see that the question of embeddability of a k-dimensional simplicial complex in $\mathbb{R}^d$ is nontrivial in general for $k \leq d \leq 2k$.

Van Kampen's and Flores's primary example was the simplicial complex consisting of all faces of dimension at most k in a $2k$-dimensional simplex. We will work with another example (also known to van Kampen and Flores).

Let D_3 denote the simplicial complex with three isolated vertices. The complex for which we prove non-embeddability in $\mathbb{R}^{2k}$ is the $(k+1)$-fold join $D_3^{*(k+1)}$. More graphically, we think of the vertex set of $D_3^{*(k+1)}$ as $k+1$ rows with 3 vertices each, and the simplices are the subsets that use at most one vertex per row. In particular, for $k = 1$ we get $K_{3,3}$, one of the two Kuratowski graphs.

An abstract version of antipodality. In order to prove non-embeddability of a simplicial complex K in $\mathbb{R}^d$, we consider a hypothetical continuous injective map $f\colon |K| \to \mathbb{R}^d$, and we want to derive a contradiction. But the injectivity condition, $f(x) \neq f(y)$ for $x \neq y$, does not appear very suitable to work with directly.

Instead, we are going to use a surprising trick of general importance: from $|K|$ and from $\mathbb{R}^d$ we construct new, more complicated-looking spaces, and the given f yields a map between these new spaces. The key advantage is that this new map satisfies a more global and convenient condition: it is antipodal in a suitable sense, and for our particular simplicial complex $D_3^{*(k+1)}$ and for $d = 2k$ we

will be able to use the Borsuk–Ulam theorem to conclude that the new map cannot exist.

The first thing to do is generalizing the definition of antipodal maps suitably. The definition of antipodal maps $\mathbb{R}^m \to \mathbb{R}^n$ given above Theorem 5.1 relies on the particular map, one for each $\mathbb{R}^n$, sending x to $-x$. The important properties are that this map is a self-homeomorphism of $\mathbb{R}^n$, and that applying it twice gives the identity.

We thus define a **$\mathbb{Z}_2$-space** as a pair (X, ν), where X is a topological space and $\nu\colon X \to X$ is a homeomorphism with $\nu\nu = \mathrm{id}_X$. (Here $\mathbb{Z}_2$ refers to the (only) two-element group $\{0, 1\}$ with addition modulo 2, indicating that our considerations could be generalized to G-spaces with G a finite group or a topological group, but we will not pursue this direction in this introductory treatment.) The homeomorphism ν is often called the *$\mathbb{Z}_2$-action* of the considered $\mathbb{Z}_2$-space.

With $\mathbb{Z}_2$-spaces as objects, we also want to define the corresponding maps (morphisms). A **$\mathbb{Z}_2$-map** between $\mathbb{Z}_2$-spaces (X, ν) and (Y, ω) is a continuous map $f\colon X \to Y$ with $f\nu = \omega f$; this is the analog of $f(-x) = -f(x)$ for antipodality.

Deleted product and the Gauss map. We return to the earlier setting with an injective continuous $f\colon |K| \to \mathbb{R}^d$. From the space $|K|$ we construct the **deleted product** $|K|^2_\Delta$, where for an arbitrary space X we have

$$X^2_\Delta := \{(x, y) : x, y \in X, x \neq y\}.$$

This is a subspace of the product $X \times X$, and the subscript Δ should suggest that we delete the *diagonal* $\Delta := \{(x, x) : x \in X\}$. Moreover, we can naturally consider X^2_Δ as $\mathbb{Z}_2$-space, with the $\mathbb{Z}_2$-action $(x, y) \mapsto (y, x)$.

Based on f, we define a $\mathbb{Z}_2$-map $\tilde{f}\colon |K|^2_\Delta \to S^{d-1}$, where the sphere on the right is considered with the usual antipodality $x \mapsto -x$, as follows:

$$\tilde{f}(x, y) := \frac{f(x) - f(y)}{\|f(x) - f(y)\|}.$$

This is sometimes called the **Gauss map**; note that it is well defined only because the deleted product contains only pairs with $x \neq y$ and because f is assumed to be injective.

We have thus arrived at the following sufficient condition for non-embeddability:

Proposition 8.2. *Let K be a finite simplicial complex, and suppose that there is no $\mathbb{Z}_2$-map $|K|^2_\Delta \to S^{d-1}$. Then $|K|$ cannot be embedded in $\mathbb{R}^d$.*

There is no obvious reason to expect the condition to be also necessary; after all, it is hard to imagine why *every* possible $\mathbb{Z}_2$-map should look like the Gauss map for some embedding f. The condition indeed may not always be necessary, but remarkably enough, for a wide range of parameters, it turns out to be necessary. This is the statement of a celebrated theorem of Haefliger and Weber, which asserts that, for $\dim K \le \frac{2}{3}d - 1$, Proposition 8.2 actually holds as equivalence: $|K|$ embeds in $\mathbb{R}^d$ if and only if a $\mathbb{Z}_2$-map $|K|^2_\Delta \to S^{d-1}$ exists. The proof, unlike for Proposition 8.2, is difficult and we will not say anything more about it.

Deleted joins. The reader might reasonably expect that we will now use Proposition 8.2 to establish the van Kampen–Flores theorem, but the problem with this is that the deleted product of our specific complexes is not so easy to work with, at least by elementary means. Instead, we first derive a variant of the proposition in which deleted *products* are replaced by deleted *joins*.

We need an analogy of the Gauss map for joins: a $\mathbb{Z}_2$-map from the twofold join X^{*2}, with something like the diagonal deleted, into a sphere, a reasonable guess being S^d instead of S^{d-1} as before, since the join has dimension one larger than the product. We recall that points of X^{*2} are triples (x, y, t), $x, y \in X$, $t \in [0, 1]$, with appropriate identifications for $t = 0$ and $t = 1$. After some experimenting, one can arrive at the following Gauss-like map formula:

$$\tilde{f}(x, y, t) := \frac{(tf(x) - (1-t)f(y), 2t - 1)}{\|(tf(x) - (1-t)f(y), 2t - 1)\|} \in S^d.$$

The expression, $(tf(x) - (1-t)f(y), 2t-1)$ on the right-hand side is a $(d+1)$-component vector with the first d components given by $tf(x) - (1-t)f(y)$ and the last one by $2t - 1$. The formula is well-defined unless both $tf(x) = (1-t)f(y)$ and $2t - 1 = 0$, or in other words, unless $f(x) = f(y)$ and $t = \frac{1}{2}$. So it is sufficient to delete all triples

$(x, x, \frac{1}{2})$ from X^{*2}, which gives us a (somewhat ad-hoc) notion of deleted join. Some care is needed to check continuity of $\tilde{f}$; we need to see that it respects the identifications in the definition of the join as a quotient space. Actually, the definition of $\tilde{f}$ has been reverse-engineered to obey these identifications. Finally, we can see that $\tilde{f}$ is a $\mathbb{Z}_2$-map if the $\mathbb{Z}_2$-action on X^{*2} is given by $(x, y, t) \mapsto (y, x, 1 - t)$, a natural choice.

Simplicial deleted join. In our setting, where $X = |K|$ is the polyhedron of a simplicial complex, the join $|K|^{*2}$ is, as we recall, the polyhedron of the simplicial complex K^{*2}. However, deleting the points $(x, x, \frac{1}{2})$ as above destroys this structure; look, for example, what happens if K is a segment (1-dimensional simplex).

We will thus define a simplicial version of the deleted join, in which we delete more points but keep a nice simplicial structure. Let K' and K'' be two vertex-disjoint copies of K, and for a simplex $F \in K$, let us write F' and F'' for the corresponding simplices in K' and K'', respectively. Then the simplicial deleted join of K is

$$K^{*2}_{\Delta} := \{F' \cup G'' : F, G \in K, F \cap G = \emptyset\}.$$

We thus delete all joins of intersecting pairs of simplices.

The polyhedron $|K^{*2}_{\Delta}|$ is again a $\mathbb{Z}_2$-space, with the same $\mathbb{Z}_2$-action as the one for the join. Moreover, it is contained, usually strictly, in the deleted join of $|K|$ as a space, as announced. Instead of a general proof, we invite the reader to work out an example or two, to see what is going on.

Exercise 8.3. Describe/visualize the deleted join of the 3-cycle and of the 4-cycle (understood as 1-dimensional complexes). Check that it does not contain any points of the usual join of the form $(x, x, \frac{1}{2})$. ✠

By the above considerations, we thus have the following analog of Proposition 8.2 with deleted joins.

Proposition 8.4. *Let K be a finite simplicial complex, and suppose that there is no $\mathbb{Z}_2$-map $|K^{*2}_{\Delta}| \to S^d$. Then $|K|$ cannot be embedded in $\mathbb{R}^d$.*

Proof of Theorem 8.1. We need to understand the deleted join $(D_3^{*(k+1)})^{*2}_{\Delta}$.

Exercise 8.5. (a) Work out what the simplicial deleted join of D_3 is (the $k = 0$ case).

(b) Check that the join and deleted join commute: $(K * L)^{*2}_{\Delta} \cong K^{*2}_{\Delta} * L^{*2}_{\Delta}$. Using this, check that $(D_3^{*(k+1)})^{*2}_{\Delta}$ is isomorphic to the join of $k+1$ copies of S^1 (represented as 6-cycles), and hence its polyhedron is an S^{2k+1}; see Exercise 6.4. ✠

Now we know that the deleted join is homeomorphic to S^{2k+1}. With some more care, it can be verified that the homeomorphism obtained in this way is also a $\mathbb{Z}_2$-map; we omit this part here, since it is not very instructive (moreover, it can be shown that every two $\mathbb{Z}_2$-actions on S^n without fixed points, i.e., with $\nu(x)$ never equal to x, are equivalent, in the sense that there is a $\mathbb{Z}_2$-map between the resulting $\mathbb{Z}_2$-spaces in both directions).

Thus, a $\mathbb{Z}_2$-map of the deleted join into S^{2k} yields an antipodal map $S^{2k+1} \to S^{2k}$, which contradicts the Borsuk–Ulam theorem (Theorem 5.1(iii)). □

9. Homotopy Groups

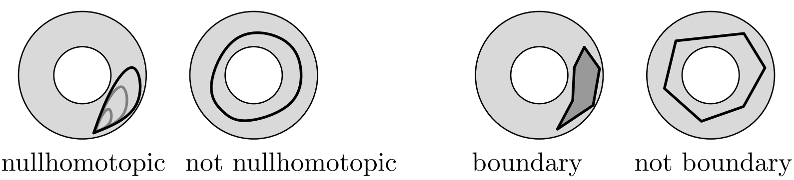
We will now introduce two fundamental concepts of algebraic topology: *homotopy groups* of a space on the one hand, and *homology groups* and the closely related *cohomology groups* on the other hand.

Rudiments of these notions go back to a theory of integration along curves in the complex plane in the 19th century. Homotopy and homology groups were explicitly introduced by Poincaré at the beginning of the 20th century, in a form somewhat different from the modern one. They were the key concepts that have made a large part of topology algebraic, in the sense of associating algebraic structures with topological spaces.

For both homotopy and homology, we can take as a starting point the question, what makes a disk topologically different from a disk with a hole, i.e., an annulus A? The first impulse may be to say that A has a hole and a disk does not, but then, what is a hole? Apparently

in the annulus case it is something outside A, and to think of such a hole, one must imagine the annulus as a subspace of something else, say the plane. But what happens to the hole if we think of A as the surface of a cylinder in $\mathbb{R}^3$, for example? We need an intrinsic notion, talking only about the space itself.

Both in homotopy and homology, we look at loops in the considered space X. From the homotopy point of view, we consider a loop as a map $S^1 \to X$. Every loop in the disk is nullhomotopic, i.e., can be continuously shrunk to a point, but the annulus has nontrivial loops, ones that are not nullhomotopic (left picture):

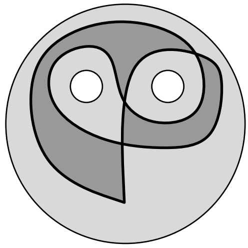


On the other hand, thinking homologically, we consider a loop just as the image of the map $S^1 \to X$, i.e., a point set,[6] and we ask, is the considered loop the boundary of something 2-dimensional[7] in X? Again, every loop is a boundary in the disk, but not so in the annulus—see the right picture (these claims should be quite intuitive, but we have not proved them).

Here is an example showing that these two notions of triviality of a loop, being nullhomotopic and being a boundary, are different. We believe that experiencing this early and hands-on is important for developing some intuition about homology later on. So now we consider a disk with two holes, and the self-intersecting loop as in the picture:

[6]This applies to a "baby" homology over a 2-element field, which we will mostly deal with in this text. In a more "grown-up" homology with integer coefficients, normally used in textbooks and in many applications, we would need to consider loops with orientation and possibly with multiplicity.

[7]Here the meaning of "boundary" is somewhat different from the definition in general topology; we want the boundary of a 2-dimensional disk in $\mathbb{R}^3$ to be a circle.



The loop is the boundary of the indicated region. But it is not nullhomotopic; again, we do not prove this, but we can perhaps recommend the reader a physical experience with a string and a bar with two sticks or something similar.

Pointed everything. We postpone further discussion of homology and focus on homotopy groups.

The idea of looking at homotopy classes of loops in the considered space still needs a refinement: we want only loops that begin and end in some distinguished point, called the *basepoint*, $x_0 \in X$, since such loops can be composed. The composition of loops a and b is the loop $a.b$ obtained by first traversing a and then b. More formally, if a is represented by a map $a\colon [0,1] \to X$ with $a(0) = a(1) = x_0$, and similarly for b, then $c = a.b$ is given by

$$c(t) = \begin{cases} a(2t) & 0 \le t \le \frac{1}{2} \\ b(2t-1) & \frac{1}{2} \le t \le 1. \end{cases}$$

In order to do everything consistently, it is good to work in the category of **pointed spaces**. A pointed space is a pair (X, x_0), where X is a topological space and $x_0 \in X$ is a basepoint. The appropriate morphisms are **pointed maps**, i.e., continuous maps sending the basepoint to the basepoint. Here are some other "pointed" notions:

- A **pointed homotopy** H of pointed maps $f, g\colon (X, x_0) \to (Y, y_0)$ is a homotopy that, moreover, fixes the basepoint all the time, i.e., $H(x_0, t) = y_0$ for all $t \in [0,1]$. Let $[f]_*$ denote the pointed homotopy class of a pointed map f (the star usually refers to the pointed setting), and let $[(X, x_0), (Y, y_0)]_*$ be the set of all such classes for given (X, x_0) and (Y, y_0).

- The **wedge** $X \vee Y$ of pointed spaces X and Y is obtained by taking the disjoint union and identifying the basepoints (and similarly for any number of spaces). This is actually the coproduct in the category of pointed spaces (while the product has an obvious unique basepoint, and so no change is needed).

- More as an interesting illustration than something we would actually need, we mention that the pointed analog of the suspension $\mathrm{S}X$ is the *reduced suspension* ΣX, obtained from $\mathrm{S}X$ by collapsing the two segments above the basepoint of X to a new basepoint (we are thinking of $\mathrm{S}X$ as a double cone over X).

Let us remark that it is very convenient and technically useful to extend pointed spaces to **pairs** (X, A), where X is a space and A is a subspace of X (usually assumed to sit nicely in X, say as a subcomplex of a finite simplicial complex). A *map of pairs* $f\colon (X, A) \to (Y, B)$ is a continuous map $X \to Y$ with $f(A) \subseteq B$—a very simple concept but with large expressive power, and a basis of notions such as *relative homotopy or homology groups*, which are important tools for working with the usual homotopy and homology groups. For example, now we can also think of a pointed map $S^1 \to X$ as a map of pairs $([0, 1], \{0, 1\}) \to (X, \{x_0\})$.

The fundamental group. We are ready to introduce the first homotopy group $\pi_1(X)$ of a pointed space, one of the great inventions of Poincaré, also called the *fundamental group*.

Definition 9.1. *Let (X, x_0) be a pointed space. Its* **fundamental group** $\pi_1(X, x_0)$ *has the ground set $[(S^1, s_0), (X, x_0)]_*$ of pointed homotopy classes of pointed loops in X, and the group operation is given by composition of loops, i.e., $[a]_*[b]_* = [a.b]_*$.*

Exercise 9.2. (a) Check that the operation is well defined.

(b) Show that it is associative (a proper picture makes this quite obvious).

(c) Take a disk with two holes and find an example of loops a, b witnessing noncommutativity of the fundamental group. The example should be just informal since we have not built the tools for showing nontriviality of any loop. ✠

The basepoints in the notation are annoying and one would like to get rid of them as soon as possible. Unfortunately, for different basepoints x_0, x_0' the groups $\pi_1(X, x_0)$ and $\pi_1(X, x_0')$ are certainly not equal (they are disjoint as sets). But, under reasonable circumstances, they are at least isomorphic, which gives us a good enough reason to ignore the basepoint.

Exercise 9.3. Let X be a space, and let x_0, x_0' be two points connected by a path γ (a map $[0,1] \to X$ with $\gamma(0) = x_0$, $\gamma(1) = x_0'$). Exhibit an isomorphism $\pi_1(X, x_0) \cong \pi_1(X, x_0')$. Hint: *The Hobbit*, full title. ✠

A space X with a trivial fundamental group (i.e., with no nontrivial loops), which is usually written as $\pi_1(X) = 0$, is called **simply connected**.[8]

Functors. From every pointed map $f\colon (X, x_0) \to (Y, y_0)$, we obtain a map $f_*\colon \pi_1(X, x_0) \to \pi_1(Y, y_0)$ of the fundamental groups: if you think of it, there is only one possible definition of f_*, namely, $f_*([a]_*) = [fa]_*$—just look at the loop's image under f. A routine check shows that this is well defined, and, moreover, one easily finds that f_* is a group homomorphism. Finally, the construction respects composition of maps: $(fg)_* = f_* g_*$.

Because of these wonderful properties, π_1 is a **functor**, more precisely, a functor from the category of pointed spaces to the category of groups. A functor is yet another key notion of category theory we wanted to mention, but we have been waiting for a good example.

A functor F from a category C to a category D assigns an object $F(X) \in \mathrm{Ob}(\mathsf{D})$ to every $X \in \mathrm{Ob}(\mathsf{C})$ and a morphism $F(f) \in \mathrm{Hom}(F(X), F(Y))$ to every morphism $f \in \mathrm{Hom}(X, Y)$ in C, so that

[8] Often this also includes the assumption that X is path-connected, but not in all sources. On the other hand, the fundamental group can "see" only the path-connected component containing the basepoint.

identity morphisms are mapped to identity morphisms and F respects composition, $F(fg) = F(f)F(g)$.

Another example of a functor we have met is the polyhedron of a finite simplicial complex, from the category of finite simplicial complexes with simplicial maps to topological space with continuous maps (or, strictly speaking, equivalence classes of homeomorphic topological spaces).

If we take a commutative diagram of pointed spaces with pointed maps and apply a functor, such as π_1, we automatically obtain another commutative diagram, in this case with groups and homomorphisms. Since many things can be expressed by commutativity of suitable diagrams, once some steps in proofs become almost mechanical, one gets used to this approach. When considering some construction taking objects of some kind and producing objects of another kind, say simplicial complexes from graphs, it may be worth asking whether it is a functor, or if it can be adjusted to behave functorially.

The preservation of commutative diagrams by functors might perhaps suggest that functors should preserve limits, such as products, but this is not the case in general—the problem is with the uniqueness requirement in the definition of limit.

On the uncomputability of the fundamental group. Most of the elementary texts on algebraic topology cover basic properties of the fundamental group and tools for working with it, such as covering spaces, as well as rigorous computations of the fundamental group of S^1 and of compact 2-dimensional surfaces. Here we do not go in this direction, mainly because in many computer-science uses of topology, spaces with a nontrivial fundamental group are difficult or impossible to deal with, while interesting things can be done once we assume a simply connected space.

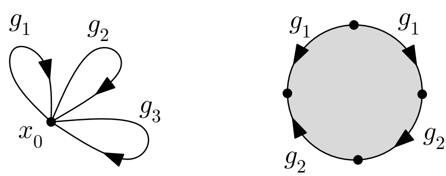
The basic difficulty with $\pi_1(X)$ is that almost nothing about it is algorithmically computable. These uncomputability results all go back to uncomputability results for groups.

A group G is said to be *finitely presented by generators and relations* if we are given a list $g_1, \ldots, g_n$ of elements of G that together generate G, plus a finite list of relations such as $g_3 g_5 g_4^7 g_2^{-3} = e$, e the

unit element of G. For example, the free group on 2 generators may have the generator list a, b and no relations, while adding the relation $aba^{-1}b^{-1} = e$ yields the abelian group $\mathbb{Z}^2$, etc.

Practically everything about groups presented in this way is algorithmically undecidable in general, such as *nontriviality* (does G have an element distinct from e?), or the *word problem* (do the defining relations of G imply a given relation?).

Given a group G finitely presented by generators and relations, one can algorithmically construct a 2-dimensional simplicial complex having G as the fundamental group. The idea is extremely simple: given the generators $g_1, \ldots, g_n$, we make a wedge of n circles, each corresponding to one generator. Then, for every relation, say $g_1 g_2^2 g_1^{-1} = e$, we take a new disk and glue its boundary to the loops, as is illustrated below, so that it does not intersect anything else from the space constructed so far:



This usually cannot be pictured in $\mathbb{R}^3$, but the resulting space is well defined. It should also be clear that it has a finite triangulation.

Basic results about the fundamental group (the Seifert–Van Kampen theorem) immediately imply that the fundamental group of the resulting space is isomorphic to G, since it has the same generators and relations. So the nontriviality of $\pi_1(K)$, already for a 2-dimensional finite simplicial complex K, is algorithmically undecidable!

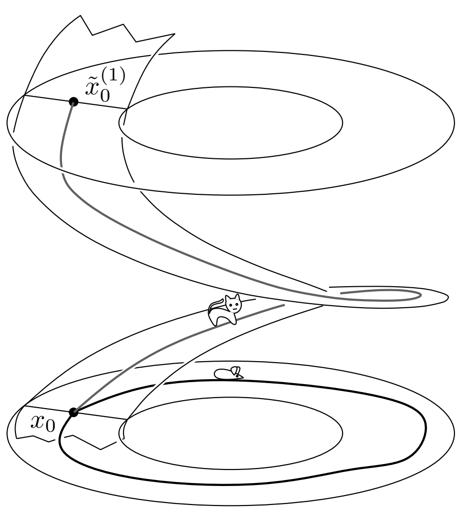
There are numerous other, more difficult undecidability results, such as the impossibility of recognizing an S^5 mentioned earlier. These usually require refinements of the basic results above both on the group side (proving uncomputability for groups of some restricted

type) and on the topological side (implanting such groups as fundamental groups in restricted classes of spaces, such as 4-dimensional manifolds).

Fortunately, while testing simple connectedness is hopeless in general, in many concrete cases it may have simple reasons and be easy to check (or be known).

So how can we prove the nontriviality of a loop? We have said that we will not develop the theory of the fundamental group, but we cannot resist sketching one of the basic tricks at least informally. We consider the annulus A with basepoint x_0 and a pointed loop a going around once, and we would like to prove that a cannot be nullhomotopic.

We think of A as a corridor in the ground (0th) floor, and we consider a spiral staircase $\tilde{A}$ that winds above A and also below it, in infinitely many loops (the geometric shape of A is called a *helicoid*; we ignore the stairs and regard $\tilde{A}$ as smooth surface). The staircase $\tilde{A}$ reaches the ground floor A exactly at the basepoint x_0, and above x_0, we have exactly one point $\tilde{x}_0^{(k)} \in \tilde{A}$ in each floor, $k \in \mathbb{Z}$, with $\tilde{x}_0^{(0)} = x_0$.



There is a mouse following the loop a in A: at time $t \in [0,1]$ it is in $a(t)$, starting and ending in x_0. A cat starts at x_0 at time 0 as well, but moves along $\tilde{A}$, not along A (we do not address a practical solution of how the mouse can freely cross $\tilde{A}$ and the cat A, but we do not let this issue distract us from the essence). The cat always stays precisely vertically above the mouse (or below it).

It seems plausible that the cat's path $\tilde{a}$ is determined uniquely (although proving this rigorously is one of the moderately technical parts we wanted to avoid), and for the particular loop a, the cat ends up at $\tilde{x}_0^{(1)}$, one floor above the mouse. Moreover, by technically very similar considerations, one can prove that if a and b are pointed-homotopic loops, then the corresponding cat's paths $\tilde{a}$ and $\tilde{b}$ are homotopic too, with homotopy fixing each of the $\tilde{x}_0^{(k)}$. In particular, the cat ends up in the same floor for both $\tilde{a}$ and $\tilde{b}$.

We can already see that our particular loop a cannot be nullhomotopic, since for the constant (unmoving) loop the cat keeps sitting in x_0. By further reasoning in this direction, one verifies that $\pi_1(A) \cong \mathbb{Z}$, the group of integers with addition.

The space $\tilde{A}$ is an example of a *covering space*, and the vertical projection $\tilde{A} \to A$ is called a *covering map*. The method of covering spaces allows one to compute, e.g., the fundamental groups of all compact 2-dimensional surfaces.

Exercise 9.4. Here is another example of a covering space: The plane $\mathbb{R}^2$ is a covering space for the torus T; we think of T as the square $[0,1]^2$ with the usual identification of the sides, and the covering map sends a point $(x,y) \in \mathbb{R}^2$ to $(x \bmod 1, y \bmod 1) \in T$. Can you deduce (informally) from this fact what the fundamental group of T is? ✠

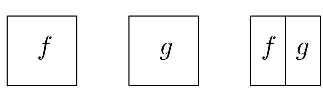
Higher homotopy groups. Higher homotopy groups $\pi_k(X)$, $k = 2, 3, \ldots$, were first introduced by Čech, long after the fundamental group $\pi_1(X)$. Čech actually withdrew his paper on the advice of senior colleagues, who believed that, unlike the groups in his definition, the true higher homotopy groups should not be commutative in general.

The definition is now the accepted one, though, and in spite of being "only" abelian, higher homotopy groups belong among the most challenging objects in mathematics. As we will see, the reason why $\pi_1(X)$ need not be commutative, but all higher homotopy groups always are, is that the 1-dimensional sphere S^1 is like a rail track where moving points cannot pass one another without colliding, while in S^k, $k \geq 2$, there is enough room for points to move around without collisions.

To define π_k, we must again consider a pointed space (X, x_0). The *elements* of $\pi_k(X)$ are easy to define with the notation we already have: they are pointed homotopy classes of pointed maps $(S^k, s_0) \to (X, x_0)$ (where s_0 is a basepoint in S^k, say the north pole).

Understanding the group operation is a bit more challenging. A good way is to regard S^k as the quotient $I^k/\partial I^k$ of the k-dimensional cube I^k, $I = [0, 1]$, by its boundary ∂I^k (we saw this representation of the sphere, with B^k instead of I^k, in Exercise 6.2). Then a pointed map $f\colon (S^k, s_0) \to (X, x_0)$ can also be regarded as a map of pairs $(I^k, \partial I^k) \to (X, x_0)$, i.e., a map from the cube that sends all of the boundary to x_0.

Let us now consider two elements $[f]_*$ and $[g]_*$ of $\pi_k(X, x_0)$, and let us think of the representing maps f, g as maps of the cube as above. Then the map h representing $[f]_* + [g]_*$ (the operation in $\pi_k(X)$, $k \geq 2$, is usually written as addition) is constructed as follows: We split I^k into the left and right half along the x_1 coordinate, we rescale the cube on which f is defined by the factor of $\frac{1}{2}$ in the x_1-coordinate and put it over the left half, and similarly the cube with g is squeezed twice and identified with the right half. Then h equals the squeezed f on the left half and the squeezed g on the right half.



Note that this directly generalizes the way we have introduced the operation in the fundamental group.

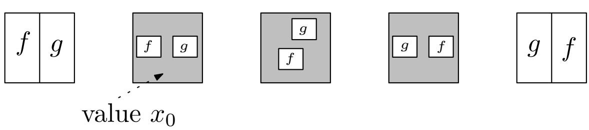
Definition 9.5. *For a pointed space* (X, x_0)*, the* ***k*-th homotopy group** $\pi_k(X, x_0)$ *is the set* $[(S^k, s_0), (X, x_0)]_*$ *of pointed homotopy equivalence classes of pointed maps of the k-sphere into* (X, x_0)*, with the addition operation described above, "putting f on the left half-cube and g on the right half-cube."*

Exercise 9.6. Describe the addition in π_2 directly using pointed maps, rather than maps of the cube. ✠

Exercise 9.7. How does one get the inverse of $[f]_*$? ✠

One again has to verify, routinely, that the operation is well defined (the result does not depend on the choice of representatives), associative, and has inverses. It is also easy to show that for a path-connected X, all choices of the basepoint x_0 give isomorphic $\pi_k(X, x_0)$. Moreover, for every $k \geq 2$, π_k is a functor from pointed spaces to abelian groups—this is not really different from the $k = 1$ case, with the exception of commutativity, which we now explain.

We want to show that the representative of $[f]_* + [g]_*$ constructed as above is homotopic to the representative of $[g]_* + [f]_*$. The homotopy, in the setting of maps from the cube, is illustrated below:



Homotopy groups of spheres. Contractible spaces, such as balls, have all homotopy groups zero. One would think that the next simplest example should be the spheres. So what is $\pi_k(S^n)$?

First, for $k < n$, it is 0. This looks quite plausible if you think of the image of S^1 in S^2, say; one should be able to pick a point of S^2 not in the image, take it as the north pole and let northern wind contract the image continuously to the south pole. There is a technical difficulty with this, since the image could be a space-filling curve covering all of S^2, but this can be dealt with, as we will see later (Corollary 11.3).

Good; now for $k = n$ we have $\pi_n(S^n) = \mathbb{Z}$ for all $n \geq 1$. This hides a nice theorem of Hopf, stating that two maps $S^n \to S^n$ are homotopic if and only if they have the same *degree*. A rigorous definition of degree can be given using homology. Here is an informal explanation: We think of S^2 in $\mathbb{R}^3$, and color it green from inside and red from outside. Then we map it to another S^2 by a map f. If f is locally sufficiently nice and we look at a generic point x of the target S^2, there are locally several sheets of the red-green S^2 over x, some of them are green when we look from inside the target S^2 and others red. The degree of f is the number of green ones minus the number of red ones. For $k = 1$ this is the winding number – number of times f "goes around" the target S^1.

How about the case $k > n$? Historically it came as a great surprise that $\pi_k(S^n)$ can be nonzero in this situation. The first instance was discovered by Hopf, who found a nontrivial map $\eta\colon S^3 \to S^2$ (i.e., not nullhomotopic). It is now called the **Hopf map** and it belongs among a handful of key examples in topology (and also with uses in quantum physics and elsewhere). The Hopf map has a one-line definition using complex numbers, but a longer explanation and pictures are needed to see what is going on—not speaking of a proof of nontriviality, which needs considerable apparatus. So we refer to the literature, e.g., [**Lyo03**].

A concise answer to the question of what the $\pi_k(S^n)$ are, after many decades of research, is—nobody knows. Many deep and interesting facts have been proved. For example, for $k > n$, the $\pi_k(S^n)$ are finite, with the sole exception of the cases $\pi_{4n-1}(S^{2n})$. One of the perhaps most remarkable phenomena is *stability*: for all $n > i+1$ the homotopy groups $\pi_{n+i}(S^n)$ depend only on i.

But these stable homotopy groups of spheres, for instance, are known only up to $i = 64$, and the known part of the table looks fairly chaotic; just to give a taste, we have $\pi_{n+35}(S^n) = \mathbb{Z}_8 \oplus \mathbb{Z}_2^2 \oplus \mathbb{Z}_{27} \oplus \mathbb{Z}_7 \oplus \mathbb{Z}_{19}$ for all $n \geq 37$.

Let us stress that these mysterious homotopy groups are, in principle, *computable*. There are algorithms that, given k and a simply connected finite simplicial complex K, output a description of $\pi_k(K)$;

some of the more recent ones might even be practically usable for k very small.

But so far such general algorithms have been useless for the problem of homotopy groups of spheres. Researchers have developed a number of very sophisticated methods tailor-made for the case of spheres; this problem has helped to keep algebraic topology progressing for several decades.

Perhaps surprisingly, most of the known $\pi_k(S^n)$ were actually computed by hand. This illustrates that, given a sufficiently structured problem and enough interest, mathematical theory can usually beat supercomputers.

10. Homology of Simplicial Complexes

Here we begin considering homology groups. They are more difficult to introduce than homotopy groups, and in some respects they carry less information, but from a computer-science perspective, they offer an immense advantage: They are efficiently computable, both in theory (polynomial-time algorithms) and in practice (reasonably large instances can be dealt with). With a bit of exaggeration one can say that in topology, whatever can be inferred from homology is efficiently computable, and almost nothing else is.

Coefficients for homology. For every integer $k \geq 0$, we want to assign an abelian group $H_k(X)$ to a topological space X, the *k-th homology group* of X. More precisely, the construction has a parameter R, which is a commutative ring (ring as in algebra); then we speak of *homology with coefficients in R* and write $H_k(X; R)$.

The most standard choice is $R = \mathbb{Z}$, the ring of integers—the notation $H_k(X)$ refers to this default case. Other common choices are $R = \mathbb{Q}$, the field of the rationals, and $R = \mathbb{Z}_p$, the finite field with p elements, p a prime. It is known that the homology groups with integer coefficients collectively encode all the information contained in homology with coefficients in any other ring R, so in a sense they are the best one can get, but other choices of coefficients may sometimes be easier to work with.

Here, in the interest of simplicity, we are going to work with coefficients in $\mathbb{Z}_2$, the two-element field. This brings two distinct kinds of simplification compared to integer coefficients.

First, we have $-1 = 1$ in $\mathbb{Z}_2$, and this relieves us from having to keep track of *orientations* of simplices and signs in formulas. Handling these issues is not conceptually difficult at all, although error-prone, but it is pleasant to ignore them at first encounter.

Second, $\mathbb{Z}_2$ is a field, and working with homology over a field is basically linear algebra—in particular, the homology groups $H_k(X; \mathbb{Z}_2)$ are really vector spaces. On the other hand, if R is not a field, then vector spaces have to be replaced with *R-modules*, a probably much less familiar concept (in particular, $\mathbb{Z}$-modules can be regarded as abelian groups).

Simplicial homology. We will follow more or less the historical route of developing homology: first, we define homology groups for a finite simplicial complex, and then we will say that the homology of a triangulable space X is the homology of a triangulation of X. For this to make sense, we must show that the result is the same for all possible triangulations, and this takes quite a bit of work.

Modern topology textbooks often prefer a different way, the so-called *singular homology.*[9] This is defined very quickly and compactly, if not too intuitively, directly for a space without a detour through triangulations, and some of the general properties are immediately obvious. However, it is difficult to compute the singular homology of almost anything, and if one wants a computational tool, some version of simplicial homology is needed anyway.

Chains, cycles, boundaries. Let K be a finite simplicial complex, and let $k \geq 0$ be an integer. Let us write K_k for the set of all simplices of dimension precisely k in K.

First we define a vector space $C_k(K; \mathbb{Z}_2)$. Its vectors are all formal linear combinations of simplices in K_k. In more detail, we fix a vector space over $\mathbb{Z}_2$ of dimension $|K_k|$, we fix a basis in it, and we identify

---

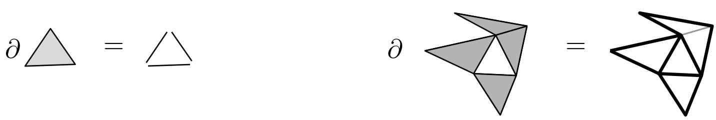
[9]There are actually many ways of defining homology groups, which may differ in generality of the considered spaces or in various fine points, so many that already in 1945 Eilenberg and Steenrod found it desirable to put all the various homology theories on an axiomatic footing, in order to isolate essential properties common to all of them.

the basis elements bijectively with the k-simplices of K. The vectors of $C_k(K;\mathbb{Z}_2)$ are called the ***k*-chains** of K, and typically they are denoted by c, or c_k if we want to stress which k we have in mind.

In the very special case of $\mathbb{Z}_2$, we can also think of k-chains as subsets of K_k: for a k-chain c, the subset consists of those simplices in K_k that have coefficient 1 in c.

Next, we define a linear map $\partial = \partial_k \colon C_k(K;\mathbb{Z}_2) \to C_{k-1}(K;\mathbb{Z}_2)$ called the **boundary operator**. It suffices to define the values of a linear map on basis vectors, and in our case we just specify the value of ∂_k on each k-simplex $F \in K_k$. Namely, $\partial_k F$ is the sum of all $(k-1)$-dimensional faces of F; formally, and recalling that we regard simplices of a simplicial complex as finite sets, we can thus write $\partial_k F = \sum_{v \in F} F \setminus \{v\}$.

The left picture illustrates this definition for $k = 2$, while the right picture shows the boundary of a more complicated 2-chain.



The boundary operator in the $\mathbb{Z}_2$ case can also easily be described combinatorially: a $(k-1)$-dimensional simplex G belongs to $\partial_k c$ for a given k-chain c exactly if it is contained in an odd number of the simplices of c.

We also make the convention that $\partial_0 = 0$; i.e., vertices have no boundary. (Sometimes it is technically convenient to change this convention slightly, which leads to the so-called *reduced homology groups* $\tilde{H}_k$, but we do not consider these here—the difference is minor anyway, and only in the 0th homology group.)

Exercise 10.1. (Trivial) To see that there is no mystery in the boundary operator, write down the matrix of ∂_2 and of ∂_1 for the simplicial complex from the right picture above. ✠

Starting from the boundary operators, we now obtain the homology groups of K by pure linear algebra. We define two vector subspaces of $C_k(K;\mathbb{Z}_2)$:

- $Z_k(K;\mathbb{Z}_2)$ consists of all k-chains whose boundary is zero (empty); concisely, $Z_k(K;\mathbb{Z}_2) = \ker \partial_k$. The vectors of $Z_k(K;\mathbb{Z}_2)$ are called ***k*-cycles** (Z is for the German word *Zyklus*).
- $B_k(K;\mathbb{Z}_2)$ consists of all boundaries of $(k+1)$-chains; i.e., $B_k(K;\mathbb{Z}_2) = \operatorname{im} \partial_{k+1}$. Its vectors are ***k*-boundaries**.

A key observation in homology is this.

Observation 10.2. *The composition $\partial_k \partial_{k+1}$ is zero (every boundary has zero boundary). Thus, $B_k(K;Z_2) \subseteq Z_k(K;\mathbb{Z}_2)$ (all boundaries are cycles).*

Proof. It suffices to prove $\partial_k \partial_{k+1} F = 0$ for every $(k+1)$-simplex F. This is immediate since every $(k-1)$-dimensional face G of F is contained in exactly two k-dimensional faces of F, namely, $F \setminus \{a\}$ and $F \setminus \{b\}$, where $F \setminus G = \{a, b\}$. □

Definition 10.3. *The k-th homology group $H_k(K;\mathbb{Z}_2)$ of a simplicial complex K is the quotient vector space $Z_k(K;\mathbb{Z}_2)/B_k(K;\mathbb{Z}_2)$, "cycles modulo boundaries."*

If we are interested only in the isomorphism type of $H_k(K;\mathbb{Z}_2)$, then a single number suffices: the dimension, which equals

$$\dim Z_k(K;\mathbb{Z}_2) - \dim B_k(K;\mathbb{Z}_2).$$

We trust that the reader can imagine how the computation of these dimensions, with K as input, could be programmed using basic subroutines for linear algebra (Gaussian elimination, say)—if not, we recommend to spend a couple of minutes on that.

For some applications, though, the isomorphism type may not suffice; we may also be interested in seeing the k-cycles representing elements of some basis of $H_k(K;\mathbb{Z}_2)$. These are also easily computed.

The elements of the quotient vector space $H_k(K;\mathbb{Z}_2)$ are equivalence classes of k-cycles, of the form $z + B_k(K;\mathbb{Z}_2)$. They are called *homology classes*. In general, two k-chains c and c' are called *homologous* if $c - c'$ is a boundary.

Here we come back to the geometric intuition sketched when we mentioned homology for the first time: the nonzero elements of the k-th homology group correspond to the k-dimensional cycles (e.g., in the 1-dimensional case, loops and linear combinations of loops) that cannot be expressed as boundaries of $(k+1)$-dimensional "things."

Enough definitions—it is time to practice a little.

Exercise 10.4. Regard a simple graph as a 1-dimensional simplicial complex. Describe k-cycles and k-boundaries in graph-theoretic terms, $k = 0, 1$. What is the meaning of H_0 and H_1? ✠

The next, very basic exercise should show that, unlike homotopy groups, the homology groups of spheres are very simple and predictable.

Exercise 10.5. (a) Consider the n-dimensional simplex as a simplicial complex (all subsets of $[n+1]$). Compute all of the homology groups with $\mathbb{Z}_2$ coefficients.

(b) Now remove the simplex itself from the simplicial complex, leaving only its boundary (this is a triangulation of S^{n-1}). What changes in the homology groups? ✠

Functors. We would like to see that, for every k, the assignment of the k-th homology group to a finite simplicial complex behaves as a functor (we still work with $\mathbb{Z}_2$ coefficients but this applies to any coefficient ring, and later on we will of course want an analog for spaces instead of simplicial complexes). Given simplicial complexes K, L and a simplicial map $f\colon K \to L$ (these are the morphisms), we want a linear map $f_{*k}\colon H_k(K;\mathbb{Z}_2) \to H_k(L;\mathbb{Z}_2)$—what should it be?

Short reflection shows that there is not much choice. First we define a linear map $f_{\#k}\colon C_k(K;\mathbb{Z}_2) \to C_k(L;\mathbb{Z}_2)$ by specifying the values on the usual basis of k-simplices. Namely, the image $f(F)$ of a simplex $F \in K$ is a simplex of L, which can be of dimension $\dim F$ or smaller. We set

$$f_{\#k}(F) := \begin{cases} f(F) & \text{if } \dim f(F) = \dim F, \\ 0 & \text{otherwise.} \end{cases}$$

Now the map f_{*k} of the homology group is defined by taking a representative k-cycle z of a homology class $h = z + B_k(K;\mathbb{Z}_2)$ and

defining $f_{*k}(h)$ as the homology class $f_{\#k}(z)+B_k(L;\mathbb{Z}_2)$. Two things could possibly go wrong with this definition; the reader is invited to check that they do not.

Exercise 10.6. (a) Check that $f_{\#(k-1)}\partial_k = \partial_k f_{\#k}$, and consequently, that the image of a cycle under $f_{\#k}$ is a cycle and the image of a boundary is a boundary.

(b) Verify that $(gf)_{*k} = g_{*k}f_{*k}$ for simplicial maps $f\colon K \to L$ and $g\colon L \to M$. ✠

Chain complexes. It is useful to isolate an intermediate algebraic object between a simplicial complex (or space) and its homology groups. A **chain complex** $\mathcal{C}$ is an infinite sequence of vector spaces (in our case where the coefficients are from a field) connected by linear maps, which are also called boundary operators:

$$\cdots \xrightarrow{\partial_3} C_2 \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} C_{-1} \xrightarrow{\partial_{-1}} \cdots$$

The sole axiom is $\partial_{k-1}\partial_k = 0$ for all $k \in \mathbb{Z}$, exactly the property of the boundary operators that allowed us to define homology groups. So we can define homology groups for an arbitrary chain complex $\mathcal{C}$ in the same way.

Of course, a primary example of a chain complex is one obtained from a simplicial complex, with $C_k = C_k(K;\mathbb{Z}_2)$ (where $C_k = 0$ for $k < 0$ by convention). But chain complexes proved useful in many other contexts, and they are the object of study of *homological algebra*.

A morphism $f_\#: \mathcal{C} \to \mathcal{D}$ of chain complexes, called a **chain map**, is modeled after the maps $f_{\#k}$ constructed above; we just postulate the property we needed so that $f_\#$ induce a map of the homology groups. Namely, $f_\# = (f_{\#k})_{k\in\mathbb{Z}}$ is a sequence of linear maps $f_{\#k}\colon C_k \to D_k$ satisfying $\partial_k f_{\#k} = f_{\#(k-1)}\partial_k$, where ∂_k on the left is the boundary operator in $\mathcal{C}$, while ∂_k on the right comes from $\mathcal{D}$.

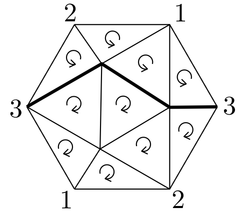
What is the point of this exercise with defining chain complexes? Besides general applicability already mentioned, there is a specific advantage for our later considerations: even between chain complexes derived from simplicial complexes, we will have chain maps that are useful and fairly obvious, but that do not come from any simplicial map.

What changes for $\mathbb{Z}$ instead of $\mathbb{Z}_2$? For integer coefficients, first of all, we must assign every simplex of K some *orientation*. A simple way is to number the vertices of K from 1 to n and then orient every simplex "from left to right". Now the definition of the boundary operator involves signs, so the boundary of a triangle is no longer the sum of its edges, but something like "the first edge minus the second edge plus the third one."

The oriented simplices appear with arbitrary integer coefficients in k-chains. The k-chains, k-cycles, and k-boundaries are no longer vector spaces, but $\mathbb{Z}$-modules or abelian groups. For computing with them we no longer suffice with linear algebra, but we need to manipulate integer matrices. Most notably, instead of Gaussian elimination, we use algorithms for computing the *Smith normal form*; these are considerably more sophisticated but reasonably well understood and fast.

The resulting homology group $H_k(K;\mathbb{Z})$ can be an arbitrary abelian group (finitely generated for a finite simplicial complex), such as $\mathbb{Z}^{17} \oplus \mathbb{Z}_2^6 \oplus \mathbb{Z}_3^2 \oplus \mathbb{Z}_8 \oplus \mathbb{Z}_{293}$. Here 17, the exponent of the infinite cyclic component, is called the *k-th Betti number* of K (warning: it need not equal $\dim H_k(K;\mathbb{Z}_2)$, the k-th $\mathbb{Z}_2$ Betti number).

The finite cyclic summands are the **torsion part** of $H_k(K;\mathbb{Z})$. They have no analog for field coefficients, and their geometric meaning is not so easy to visualize. The simplest case can perhaps be shown for the projective plane, which can be obtained by taking a disk and identifying every pair of diametrically opposite points.



The picture shows a triangulation of the projective plane (the smallest one, actually; identifications of vertices are indicated by the numbers 1,2,3) with some chosen orientations of the triangles. The

curve drawn thick is a closed loop, a homological 1-cycle (zero boundary). The curve itself is not a boundary (with $\mathbb{Z}$ coefficients), but if we take it *twice*, it becomes a boundary of the sum of all 2-simplices of the projective plane with the given orientations; note that the orientations of adjacent triangles flip locally exactly along our curve.

Cohomology. First we recall from linear algebra that if V is a vector space over a field $\mathbb{K}$, then the set of all linear functions $f\colon V \to \mathbb{K}$ also forms a vector space V^*, called the *dual* of V. If V is finite-dimensional, then $\dim V^* = \dim V$; given a basis $b_1, \ldots, b_n$ of V, we can form a basis $b_1^*, \ldots, b_n^*$ of V^*, where b_i^* attains value 1 on b_i and value 0 on all the other b_j.

Every linear map $L\colon V \to W$ determines a linear map $L^*\colon W^* \to V^*$ (note the change of direction; this will be happening all the time here), called the *adjoint* of L. There is only one reasonable way of defining L^*, namely, $L^*(f) := fL\colon V \to \mathbb{K}$, $f \in W^*$.

To introduce cohomology with $\mathbb{Z}_2$ coefficients (or with any field coefficients, there is no difference), we apply this kind of duality to the vector spaces $C_k(K;\mathbb{Z}_2)$ of k-chains and the boundary operators ∂_k. So $C^k(K;\mathbb{Z}_2)$, the k-*cochains*, is the vector space dual to $C_k(K;\mathbb{Z}_2)$.

How should one think of k-cochains? It is enough to specify the values of a k-cochain on a basis, in this case the standard basis of the k-chains consisting of k-simplices—so a k-cochain "looks" precisely like a k-chain.[10]

Next, the **coboundary operator** $\delta_k\colon C^k(K;\mathbb{Z}_2) \to C^{k+1}(K;\mathbb{Z}_2)$ now *increases* dimension by 1, and it is the adjoint of ∂_{k+1}. The k-cocycles $Z^k(K;\mathbb{Z}_2) := \ker \delta_k$ and the k-coboundaries $B^k(K;\mathbb{Z}_2) := \operatorname{im} \delta_{k-1}$ are defined analogously to the case of homology, and the k-th **cohomology group** is then

$$H^k(K;\mathbb{Z}_2) := Z^k(K;\mathbb{Z}_2)/B^k(K;\mathbb{Z}_2).$$

Exercise 10.7. Similar to homology, interpret the coboundary operator, cocycles, and coboundaries in graph-theoretic terms (for 1-dimensional simplicial complexes). ✠

[10]In the finite-dimensional case, that is. For infinite simplicial complexes, say, a k-chain may attain only finitely many nonzero values, while a k-cochain may have any number of nonzeros.

It turns out that cohomology groups are similar to homology groups in many respects, and in particular, all cohomology groups of a space X can be deduced from the knowledge of all homology groups of X. So why bother with cohomology?

Perhaps the key reason is that if we look at a simplicial map $f\colon K \to L$ (or, later, a continuous map), the induced map in cohomology goes in the *opposite direction*, $f^{*k}\colon H^k(L;\mathbb{Z}_2) \to H^k(K;\mathbb{Z}_2)$. This is because of the dual nature of cohomology, and the reader may want to contemplate the definition of f^{*k} for a while.

Category theorist say that cohomology groups are **contravariant functors**; all of the functors mentioned so far were **covariant**. This is often useful, since cohomology can be pulled in the direction opposite to homology (and homotopy groups).

One of the most important manifestations of this general phenomenon is the existence of a product structure on the union of all cohomology groups of a given space. Namely, one can define the **cup product** in cohomology (we will not do this here, because this would take us too far), which assigns to two cohomology classes $c^p \in H^p(X)$ and $c^q \in H^q(X)$ a cohomology class $c^p \smile c^q \in H^{p+q}(X)$. This operation makes the direct sum $\bigoplus_{k=0}^{\infty} H^k(X)$ into a ring, the *cohomology ring* of X. For homology this does not work in general, since one of the key maps needed for defining a product goes in the wrong direction.

The cohomology ring in general carries strictly more information about X than just all homology groups or all cohomology groups. The cup product is also used for formalizing intuitive geometric notions such as *linking number* (given two images of S^1 in $\mathbb{R}^3$, how many times are they "linked"?, or similarly for an image of S^p and an image of S^q in $\mathbb{R}^{p+q+1}$).

11. Simplicial Approximation

We have assigned a sequence of homology groups to a finite simplicial complex. To make the construction topologically useful, we still need to show that the homology groups depend only on the considered

topological space (and actually, only on its homotopy type) and not on the chosen triangulation.

We will present a substantial part of a proof. Along the way, we will encounter a surprising phenomenon concerning triangulations. As we have mentioned earlier, the usual contemporary proof would go via singular homology. However, the considerations needed to connect singular homology to simplicial homology are also nontrivial.

In the present section we will not talk about homology, but we prepare a tool for approximating continuous maps by simplicial maps—the simplicial approximation theorem. Historically, this kind of statement made possible the first rigorous (and correct, or almost correct) proofs of foundational results such as that $\mathbb{R}^m$ is not homeomorphic to $\mathbb{R}^n$ for $m \neq n$.

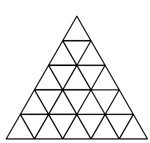
Subdivisions, especially barycentric. A geometric simplicial complex Δ' is called a **subdivision** of a geometric simplicial complex Δ if their polyhedra are the same, $|\Delta'| = |\Delta|$, and every simplex of Δ' is fully contained in some simplex of Δ.

The idea of subdivision is that each simplex of Δ is further sliced into smaller simplices, but this must be done so that the slicings are compatible across lower-dimensional faces.

The main thing about subdivisions one usually needs is that every finite geometric simplicial complex has arbitrarily fine subdivisions:

Lemma 11.1. *Let Δ be a (finite) geometric simplicial complex, considered as a subspace of $\mathbb{R}^n$ with the Euclidean metric, and let $\varepsilon > 0$ be given. Then there is a subdivision Δ' of Δ in which no simplex has diameter larger than ε.*

A two-dimensional picture



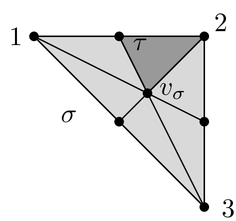
might mislead one into considering the lemma trivial. But the "parallel layer slicing" idea does not work in dimensions 3 and higher. The

usual proof of the lemma, shown below, uses a subdivision which is not very economical for the purpose of reducing the diameter of the simplices, but very simple to describe.

Sketch of proof. First we define the **barycentric subdivision** of a simplicial complex Δ. We formulate the definition recursively.

If $\dim \Delta = k$, we first produce the barycentric subdivision, call it Γ, of the subcomplex of Δ consisting of all simplices of dimension at most $k-1$ (by induction on k; for $k=0$ we do nothing). Then, for every k-dimensional simplex σ of Δ, we consider the barycenter (center of gravity) v_σ of σ, and for every simplex $\tau \in \Gamma$ lying on the boundary of σ, we construct the cone with base τ and apex v_σ (which is a simplex). All of these simplices together with those of Γ form the barycentric subdivision of Δ.

The picture illustrates the procedure for a single 2-dimensional simplex:



Let us mention in passing that there is also a direct, combinatorial description: simplices of the barycentric subdivision of Δ are in one-to-one correspondence with chains of nonempty simplices of Δ under inclusion. For example, the darker 2-simplex in the picture above corresponds to the chain $\{2\} \subset \{1,2\} \subset \{1,2,3\}$. This correspondence has several nice uses, but we will consider barycentric subdivisions only for the purposes of the present proof.

For us it is important that if we apply the barycentric subdivision to a simplex of diameter at most 1, then each simplex in the subdivision has diameter at most $1 - \frac{1}{n+1}$ (this is not so hard to verify, but we will accept it as a fact without proof). Hence if we start with a given complex Δ and iterate the barycentric subdivision sufficiently

many times, we are guaranteed to get the diameter of all simplices as small as needed. □

Now we state the main theorem of this section. Exceptionally, it contains an undefined term "simplicial approximation"; we prefer to develop the definition later. Even ignoring this term the theorem still contains a highly interesting statement.

Theorem 11.2 (Simplicial approximation theorem)**.** *Let $f\colon |K| \to |L|$ be an arbitrary continuous map of polyhedra of two finite simplicial complexes. Then there is a subdivision K' of K and a simplicial approximation $s\colon V(K') \to V(L)$ of f. In particular, s is a simplicial map such that the affine extension $|s|\colon |K'| \to |L|$ is homotopic to f.*

Note that only the "source" complex K is subdivided, while L stays as is. The map $|s|$ need not be close to f in metric sense; we are only guaranteed that they are homotopic.

Here is a nontrivial consequence:

Corollary 11.3. *Every continuous map $f\colon S^{n-1} \to S^n$ is nullhomotopic.*

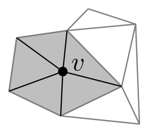
Proof. If we have a point y not in the image of f, the argument is obvious: we continuously push the image of f away from y until all of it ends up in the opposite point $-y$.

But the catch is that f may be surjective, like a space-filling curve. Then the simplicial approximation theorem comes to rescue: We consider S^{n-1} and S^n as polyhedra of K and L, respectively, and obtain a simplicial map $|s|$ from a subdivision of K into L that is homotopic to f. Such an $|s|$ cannot be surjective, because its image in L consists of simplices of dimension at most $n-1$. □

Elementary homotopies and stars. We still need to define the meaning of "simplicial approximation." First we formulate a simple condition for two continuous maps $f, g\colon |K| \to |L|$ to be homotopic. Namely, if we assume "for every $x \in K$, $f(x)$ and $g(x)$ share a simplex of L," then surely $f \sim g$, since then the desired homotopy can be

obtained by moving $f(x)$ towards $g(x)$ along a segment within the simplex containing both of these points. Let us say that f and g satisfy the *elementary homotopy condition.*

Next, we state a rather different-looking condition. Let K be a simplicial complex and let v be a vertex of it. The **open star** $\mathrm{st}^\circ v$ in K is defined as $|K|$ minus the union of all simplices of K that do not contain v.[11]



Let $f\colon |K| \to |L|$ be a continuous map of polyhedra and let $s\colon V(K) \to V(L)$ be a map of the vertex sets of the underlying simplicial complexes; at this moment we (exceptionally) do not a priori assume s simplicial. We say that f and s satisfy the *star condition* if for every vertex $v \in V(K)$ we have

$$f(\mathrm{st}^\circ v) \subseteq \mathrm{st}^\circ(s(v)).$$

Interestingly, the elementary homotopy condition and the star condition turn out to be equivalent, and if one of them holds (and hence both) for a continuous map $f\colon |K| \to |L|$ and a simplicial map $s\colon V(K') \to V(L)$, we say that s is a **simplicial approximation** of f.

Exercise 11.4. (a) Assuming that f and s satisfy the star condition, prove that s is necessarily simplicial.

(b) Show that if f and s satisfy the star condition, then f and $|s|$ satisfy the elementary homotopy condition.

(c) (Optional) Prove that, conversely, if f and $|s|$ satisfy the elementary homotopy condition, where s is simplicial, then f and s satisfy the star condition. ✠

A simple but very useful observation is that simplicial approximations respect composition: if $f\colon |K| \to |L|$ and $g\colon |L| \to |M|$ are

[11] The standard definition in the literature looks different, but from our version it is clear that the open star is an open set, which is far from clear for the standard formulation.

continuous, $s\colon V(K) \to V(L)$ is a simplicial approximation of f, and $t\colon V(L) \to V(M)$ is a simplicial approximation of g, then ts is a simplicial approximation of gf.

Exercise 11.5. [Lebesgue covering lemma] Let X be a compact metric space, and let $\mathcal{U}$ be an open cover of X. Prove that there exists $\delta > 0$ (the *Lebesgue number* of the covering $\mathcal{U}$) such that for every $x \in X$, there is $U \in \mathcal{U}$ that contains x together with its open δ-neighborhood. ✠

Proof of Theorem 11.2. Given a map $f\colon |K| \to |L|$ as in the theorem, we consider the open cover $\mathcal{U}$ of $|K|$ with preimages of open stars in L: $\mathcal{U} := \{U_w = f^{-1}(\mathrm{st}^\circ w) : w \in V(L)\}$. We let $\delta > 0$ be a Lebesgue number of this cover, and let K' be a refinement of K in which simplices have diameter below δ.

We want to construct s so that f and s satisfy the star condition. For a vertex $v \in V(K)$, the open star $\mathrm{st}^\circ v$ is contained in the δ-neighborhood of v, and hence in some $U = U_w \in \mathcal{U}$. So $f(\mathrm{st}^\circ v) \subseteq \mathrm{st}^\circ w$. It suffices to put $s(v) := w$ (choosing arbitrarily if there are several possibilities); then f and s satisfy the star condition. □

12. Homology Does Not Depend on Triangulation

In the subsequent considerations we stick to the $\mathbb{Z}_2$ coefficients in the notation, but the arguments work for any coefficients.

The case of a subdivision. First, we consider the setting where K is a finite simplicial complex and K' is a subdivision of it. As expected, there is an isomorphism $H_k(K;\mathbb{Z}_2) \cong H_k(K';\mathbb{Z}_2)$ (for every k), but this is not completely easy to prove.

Let us describe the maps inducing this isomorphism and its inverse. In the direction $K \to K'$, there seems to be no reasonable simplicial map to use, but there is a very natural chain map. Namely, for every k, we let $\lambda_{\#k}\colon C_k(K;\mathbb{Z}_2) \to C_k(K';\mathbb{Z}_2)$ be the linear map that assigns to every k-simplex σ of K the sum of all simplices $\tau \in K'$ that are contained in σ. It is easy to check that this is indeed a chain map, and so it induces homomorphisms λ_{*k} in homology.

For the other direction, we use a simplicial map $\gamma\colon V(K') \to V(K)$. Namely, as in the proof of the simplicial approximation theorem (Theorem 11.2), we obtain γ as a simplicial approximation to the identity $\mathrm{id}\colon |K'| \to |K|$. For this, it suffices to observe that whenever $v \in V(K')$ is a vertex, we can find a vertex $w \in V(K)$ with $\mathrm{st}^\circ v \subseteq \mathrm{st}^\circ w$; then we set $\gamma(v) := w$ and the star condition holds.

We leave the following result without proof (referring, e.g., to Munkres [**Mun84**]).

Fact 12.1. *The homomorphisms* $\lambda_{*k}\colon H_k(K;\mathbb{Z}_2) \to H_k(K';\mathbb{Z}_2)$ *and* $\gamma_{*k}\colon H_k(K';\mathbb{Z}_2) \to H_k(K;\mathbb{Z}_2)$ *are mutually inverse, and thus isomorphisms.*

The proof uses two additional tools, *chain homotopy* and the method of *acyclic carriers*, which we prefer not to discuss here for space reasons. Interestingly, the proof is easier if K' is the barycentric subdivision of K. This case is actually sufficient for the proof that homology does not depend on triangulation, since it is enough to have arbitrarily fine subdivisions, but their structure does not matter.

Later we will need the following (easier) statement, which is proved using chain homotopy as well, and whose proof we also omit.

Fact 12.2. *Let* $f\colon |K| \to |L|$ *be a continuous map of polyhedra and let* $s, t\colon V(K) \to V(L)$ *be two simplicial approximations of* f. *Then they induce the same homomorphism in homology:* $s_* = t_*$.

(In accordance with our earlier notation, we should write s_{*k} and t_{*k}, but since all of the considerations are valid for every k, from now on we will mostly use the simpler notation.)

The Hauptvermutung. Once topologists proved that a simplicial complex and a subdivision of it have isomorphic homology groups, they hoped to obtain the independence of homology of the triangulation by establishing the following conjecture, which became known as the **Hauptvermutung** (main conjecture, from 1908): *Every two triangulations of a triangulable topological space have a common refinement.*

However, the conjecture was much more resistant to attacks than expected, and after some time researchers found a way of bypassing

it (which will be presented below). Much later it turned out that the difficulty in proving the conjecture has a good reason: as shown by Milnor in 1961, the Hauptvermutung is actually *false*.

By now there are many examples and a great related theory concerned mainly with triangulations of manifolds (the Kirby–Siebenmann classification of manifolds and related developments), but the failure of the Hauptvermutung remains one of the truly mind-boggling phenomena in geometric topology. (See, e.g., Rudyak [**Rud01**] for a somewhat advanced exposition; unfortunately, we are not aware of any treatment easily accessible to beginners.)

Functorial properties. In order to bypass the Hauptvermutung, we will consider the homomorphisms induced by continuous maps in homology.

For a continuous map $f\colon |K| \to |L|$, we can define the induced homomorphism f_* in homology using the simplicial approximation theorem. Namely, we choose a subdivision K' of K such that there is a simplicial approximation $s\colon V(K') \to V(L)$ of f, and we set

$$f_* := s_*\lambda_* = s_*\gamma_*^{-1},$$

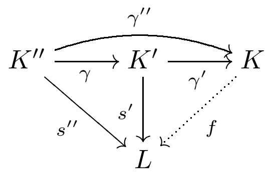
where λ_* and γ_* are the mutually inverse isomorphisms of $H_k(K;\mathbb{Z}_2)$ and $H_k(K';\mathbb{Z}_2)$ as in Fact 12.1.

The next complicated-looking statement tells us that this definition of f_* does not depend on K'.

Lemma 12.3. *Let K' and K'' be subdivisions of K, let $\gamma'\colon V(K') \to V(K)$ and $\gamma''\colon V(K'') \to V(K)$ be simplicial approximations of the identity, and let s', s'' be simplicial approximations of $f\colon |K| \to |L|$, where s' is defined on K' and s'' on K''. Finally, let $f'_* = s'_*(\gamma'_*)^{-1}$ and $f''_* = s''_*(\gamma''_*)^{-1}$ be the homomorphisms in homology as above. Then $f'_* = f''_*$.*

Proof. First we assume that there is a simplicial approximation $\gamma\colon V(K'') \to V(K')$ of the identity map $\mathrm{id}\colon |K''| \to |K'|$ (this does not mean that K'' is a refinement of K'!); later we will see how to

arrange this. Here is a diagram summarizing the situation:



We use Fact 12.2 twice: first, $s'\gamma$ and s'' are both simplicial approximations of f defined on K'', and so $s'_*\gamma_* = s''_*$, and second, $\gamma'\gamma$ and γ'' are both simplicial approximations of the identity defined on K'', and thus $\gamma'_*\gamma_* = \gamma''_*$.

Then we calculate quite mechanically

$$\begin{aligned} f''_* &= s''_*(\gamma''_*)^{-1} = s'_*\gamma_*(\gamma''_*)^{-1} = s'_*\gamma_*(\gamma'_*\gamma_*)^{-1} \\ &= s'_*\gamma_*(\gamma_*)^{-1}(\gamma'_*)^{-1} = s'_*(\gamma'_*)^{-1} = f'_*. \end{aligned}$$

Now in the general case, with K' and K'' arbitrary, we fix a subdivision K''' such that the identity has simplicial approximations $\gamma_1 \colon V(K''') \to V(K')$ and $\gamma_2 \colon V(K''') \to V(K'')$. Then we use the result of the first part twice, once on $K''' \to K' \to K$ and once on $K''' \to K'' \to K$. □

Exercise 12.4. Do the second step (the general case) in the above proof carefully in detail. How exactly do we get K'''? ✠

Theorem 12.5 (Functoriality of homology)**.**

(i) *Let* $\mathrm{id} \colon |K| \to |K|$ *be the identity map. Then the induced homomorphism in homology* id_* *is the identity.*

(ii) *Let* $f \colon |K| \to |L|$ *and* $g \colon |L| \to |M|$ *be continuous maps of polyhedra. Then* $(gf)_* = g_* f_*$.

Part (i) is immediate if we use the identity simplicial map $V(K) \to V(K)$ in the definition of id_*. Part (ii) is proved in a way very similar to the proof of the previous lemma, again using Fact 12.2 twice.

Exercise 12.6. Prove the theorem. Draw a suitable diagram of the maps involved. ✠

The theorem immediately implies the result we wanted to establish.

> Two triangulations of the same space give the same homology groups. A homeomorphism of triangulable spaces induces isomorphism in homology.

Indeed, if $h\colon |K| \to |L|$ is a homeomorphism, then we have $h_*(h^{-1})_* = (hh^{-1})_* = \mathrm{id}$, and so h_* has an inverse.

With some more work, one can show that homotopic maps induce the same homomorphism in homology: $f \sim g$ implies $f_* = g_*$ (the main idea is to use a simplicial approximation of the homotopy between f and g). It follows that not only homeomorphic, but also homotopy equivalent spaces have the same homology.

13. A Quick Harvest and Two More Theorems

The sphere S^n has nonzero homology groups exactly in dimensions 0 and n. Hence S^n is not homotopy equivalent to S^m for $m \neq n$.

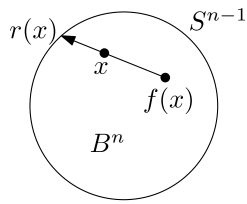
Corollary 13.1 (Invariance of domain)**.** *$\mathbb{R}^m \not\cong \mathbb{R}^n$ for $m \neq n$.*

This result may look innocent, but finding a rigorous proof had been elusive for several decades, until Brouwer found a (very complicated) solution in 1910.

Proof. A homeomorphism $\mathbb{R}^m \cong \mathbb{R}^n$ would yield a homeomorphism $\mathbb{R}^m \setminus \{0\} \cong \mathbb{R}^n \setminus \{0\}$. But $\mathbb{R}^m \setminus \{0\}$ is homotopy equivalent to S^{m-1} and $\mathbb{R}^n \setminus \{0\}$ to S^{n-1}. □

> **Theorem 13.2** (Brouwer's fixed-point theorem)**.** *Every continuous map $f\colon B^n \to B^n$ has a fixed point; i.e., there exists $x \in B^n$ with $f(x) = x$.*

Proof. The first step appears in almost all proofs of Brouwer's theorem. Assuming for contradiction that there is an f with $f(x) \neq x$ for all x, we construct a map $r\colon B^n \to S^{n-1}$ as in the picture:



That is, we send a ray from $f(x)$ to x, and $r(x)$ is the point where it hits the boundary of the ball.

The (hypothetical) map r is a *retraction* of the ball onto the sphere, meaning that $r(B^n) = S^{n-1}$ and on S^{n-1}, r acts as the identity. Using homology, we will show that r cannot exist.

Let $i\colon S^{n-1} \to B^n$ stand for the inclusion map. The composition ri is the identity $\mathrm{id}_{S^{n-1}}$, so we have the following commutative diagram:

$$S^{n-1} \xrightarrow[i]{} B^n \xrightarrow[r]{} S^{n-1} \qquad (\text{composite} = \mathrm{id})$$

Now we apply the $(n-1)$st homology group functor $H_{n-1}(.;\mathbb{Z}_2)$ and obtain

$$H_{n-1}(S^{n-1};\mathbb{Z}_2) \xrightarrow[i*]{} H_{n-1}(B^n;\mathbb{Z}_2) \xrightarrow[r_*]{} H_{n-1}(S^{n-1};\mathbb{Z}_2). \qquad (\text{composite} = \mathrm{id})$$

Substituting for the homology groups involved yields

$$\mathbb{Z}_2 \longrightarrow 0 \longrightarrow \mathbb{Z}_2. \qquad (\text{composite: } \cong)$$

But this is impossible, since an isomorphism $\mathbb{Z}_2 \cong \mathbb{Z}_2$ cannot factor through the trivial group. □

Two nice theorems. We conclude our discussion of homology by presenting two more advanced results, which may be useful to know about.

The first one is a theorem of Hurewicz relating homology and homotopy groups. We state a somewhat special case of the theorem,

which may be easier to grasp and remember than the usual general statement.

Theorem 13.3 (Hurewicz)**.** *If X is a simply connected space (i.e., $\pi_1(X) = 0$), then the first nonzero homotopy group and the first nonzero integral homology group occur in the same dimension and they are isomorphic. That is, for some $k \geq 2$ (possibly $k = \infty$) we have $\pi_j(X) = H_j(X;\mathbb{Z}) = 0$ for all $j < k$, and $\pi_k(X) \cong H_k(X;\mathbb{Z}) \neq 0$.*

Moreover, for every space X, $H_1(X;\mathbb{Z})$ is the abelianization of $\pi_1(X)$, i.e., the quotient of $\pi_1(X)$ by its commutator subgroup.

The second result we want to highlight is concerned with a situation where we have the sphere S^n and a nonempty proper closed subspace $X \subset S^n$, and we would like to say something about the topological properties of the complement $S^n \setminus X$.

The unit circle S^1 can be embedded in S^3 in the "standard" way, say as a great circle, but also in various knotted ways, for example as the trefoil knot. It can be shown that the complements of the unknotted circle and of the trefoil are not homeomorphic (for instance, they have nonisomorphic fundamental groups). Therefore, the topology of X does not determine the topology of the complement.

Surprisingly, though, all *homology groups* of $S^n \setminus X$ are determined by X, and actually by the cohomology groups of X—this is the claim of a theorem known as the Alexander duality.

We should stress that our definition of homology using finite simplicial complexes is not adequate for this setting, since even if X is triangulable by a finite simplicial complex, its complement is an open set, which does not admit a finite triangulation.

Thus, one has to use other, more general definitions of homology. With the most usual one, *singular (co)homology*, we must make an extra assumption on X, namely, that X be *locally contractible* (every point has a contractible neighborhood). To get a general statement, one uses yet another definition, the *Čech cohomology*.[12]

[12] All of these definitions agree with the simplicial one in the case of triangulable spaces, and they differ for quite pathological sets.

The last remark before the statement of the Alexander duality is that we need to use *reduced* homology and cohomology, which is marked by a tilde above H. It influences only the 0th (co)homology group: while $H_0(X;\mathbb{Z})$ is $\mathbb{Z}^c$, where c is the number of path-connected components of X, $\tilde{H}_0(X;\mathbb{Z})$ is $\mathbb{Z}^{c-1}$.

Theorem 13.4 (Alexander duality)**.** *Let $X \subset S^n$ be a nonempty proper closed subset of the sphere. Then we have, for $i = 0, 1, \ldots, n-1$,*

$$\tilde{H}^{n-i-1}(X;\mathbb{Z}) \cong \tilde{H}_i(S^n \setminus X;\mathbb{Z}),$$

where we need to use Čech cohomology on the left-hand side in general, and singular homology on the right-hand side.

For example, the Alexander duality implies that the complement of S^1 embedded in S^2 has two connected components, which is the contents of the famous *Jordan curve theorem.*

14. Manifolds

Manifolds constitute the most studied and most often applied class of topological spaces. Motivation and examples, besides pure mathematics, come mainly from physics, where certain kinds of manifolds play an essential role.

> A second-countable Hausdorff topological space M is a **manifold** if for every point $x \in M$ there exists an open neighborhood U_x of x in M that is homeomorphic to some $\mathbb{R}^n$.

Simple examples of manifolds are R^n, S^n, the torus, or the projective plane. A more challenging type of example is $\mathrm{SO}(\mathbb{R}, n)$, the group of all rotations in $\mathbb{R}^n$ around the origin, which can be represented by the set of all orthogonal $n \times n$ real matrices with unit determinant. Here, the condition in the definition of a manifold is non-obvious.

The number n in the definition above has to be constant for all points in every path-connected component of M, and it is called the **dimension** of that component. Some authors insist that all components have the same dimension as well (in this case we speak of an

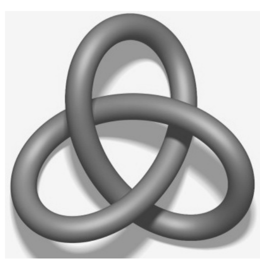
n-manifold and sometimes write M^n), while others admit combining components of different dimensions, but this is just a matter of convention.

The assumption "second-countable and Hausdorff" is made to exclude pathologies such as the long ray of Example (F) in Section 2. Actually, most texts focus on *compact* manifolds, where a number of annoying technical difficulties disappear, and we will follow the suit. As examples of non-compact manifolds, one may think of an infinite cylinder or an infinite string of tori.

A **manifold with boundary** is defined in almost the same way as a manifold, only the neighborhood U_x is allowed to be homeomorphic *either* to $\mathbb{R}^n$ *or* to a closed halfspace in $\mathbb{R}^n$.

The points with a neighborhood homeomorphic to $\mathbb{R}^n$ are called *interior points*, and the remaining points form the *boundary*.

An obvious example of a manifold with boundary is the ball B^n. For a more interesting one, let us consider a knot in S^3. Mathematically, a knot is an S^1 embedded in S^3, but fancier drawings of knots show a thickened S^1: we imagine that we drill a thin non-self-intersecting tunnel in S^3 along the embedded S^1.



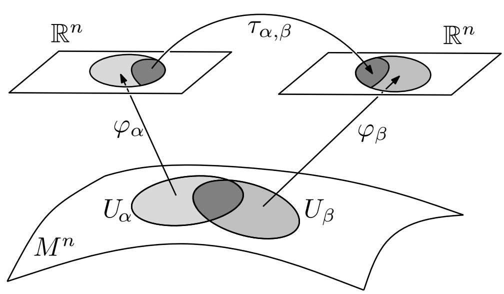
If we consider the tunnel T as an open set, so that $S^3 \setminus T$ is closed, then $S^3 \setminus T$ is a 3-manifold with boundary called a *knot manifold*.

Atlases and additional structures on manifolds. The manifolds we have defined above are **topological manifolds**, i.e., topological spaces satisfying an additional condition. But very often one wants a manifold to carry more structure, in order to calculate derivatives of functions in given directions, tangent spaces, curvature, and similar quantities, to integrate real functions or more complicated

quantities, to set up and solve differential equations, or to do all kinds of other mathematics. Indeed, historically manifolds have emerged as an abstraction of such situations in various areas, most notably in the geometry of surfaces, theory of elliptic integrals, complex analysis, and analytical mechanics.

A device for introducing such additional structures on a manifold is an **atlas**. An atlas for a manifold M^n is a collection of pairs $(U_\alpha, \varphi_\alpha)_{\alpha\in\Lambda}$, where each U_α is an open subset of M^n, and $\varphi_\alpha\colon U_\alpha \to \mathbb{R}^n$ is a homeomorphism of the set U_α with an open subset of $\mathbb{R}^n$. Unlike in atlases of Earth, the pairs $(U_\alpha, \varphi_\alpha)$ are called *charts*, not maps,[13] but otherwise, an atlas of Earth gives a good example. All charts together are required to cover all of M^n.

An important concept is a *transition map*. It is a map that answers the question, If I am now here on my map, am I also on your map, and where? More formally, if $U_\alpha \cap U_\beta \neq \emptyset$, the transition map $\tau_{\alpha,\beta}$ is the composition $\varphi_\beta\varphi_\alpha^{-1}$, which goes from $\varphi_\alpha(U_\alpha \cap U_\beta)$ to $\varphi_\beta(U_\alpha \cap U_\beta)$, both open sets in $\mathbb{R}^n$.



An atlas is called *differentiable* (or *smooth*) if all of the transition maps are of the class C^∞, i.e., if they have continuous partial derivatives of all orders (this is well defined since the transition maps are maps between open subsets of $\mathbb{R}^n$). A **differentiable** or **smooth manifold** is a manifold equipped with a differentiable atlas. (One also says that the manifold has a *differentiable structure*.)

[13] Or actually, sometimes the φ_α are also called *coordinate maps*.

Informally, one can think of the differentiablity of an atlas as follows: a path that is drawn as a smooth curve on one of the charts must be smooth on all other charts.

The point of this definition is that on such a manifold, we can define which real functions are differentiable, and develop a (coordinate-free) calculus for them (the latter belongs to differential geometry and we will not discuss it here). Namely, a function $f\colon M^n \to \mathbb{R}$ is differentiable if each of the functions $f\varphi_\alpha^{-1}$ is differentiable (C^∞) in the usual sense. The assumed differentiability of the transition maps guarantees that this definition is globally consistent.

In a similar way, one could define a C^k-atlas for given k; we recall that C^k is the class of functions with continuous partial derivatives up to order k. However, these are seldom mentioned in the literature.

The reason is a theorem of Whitney, stating that once a manifold has a C^k structure S, there is a unique C^∞ structure S' on it that is equivalent, as a C^k-structure, to S. So in this sense, there is only one notion of differentiable manifold.

What differentiable structures are possible on the most basic manifolds like $\mathbb{R}^n$ or S^n? It seems hard to imagine that there could be any others but the standard ones, but people studying this question discovered quite strange phenomena.

As for $\mathbb{R}^n$, the differentiable structure is unique for all n *except* for $n = 4$, where there are uncountably many mutually nonequivalent differentiable structures.

The first *exotic sphere*, i.e., a sphere with differentiable structure not equivalent to the standard one, was discovered by Milnor in dimension 7, and later it was found that S^7 admits 28 nonequivalent structures. The general picture is quite complicated and it is, among others, related to the stable homotopy groups of spheres; in some dimensions there is just one structure, and typically there are finitely many.

The recipe for defining differentiable manifold is quite general. Once we require that the transition maps of an atlas belong to some particular class of maps, we can say that the corresponding manifold

is of that class. The literature abounds with such classes; e.g., *analytic manifolds*, *complex manifolds*, *symplectic manifolds*, or *contact manifolds*.

Another important class, whose definition does not quite fit the pattern above, though, are *Riemannian manifolds*—see, e.g., the book [**Ber03**] for interesting views of Riemannian geometry, which we neglect here completely.

Two, three, four, many. The difficulty of studying manifolds depends very much on their dimension. And surprisingly, the dependence is not "the higher dimension, the harder" as one might perhaps expect—it is four-dimensional manifolds that pose the most tantalizing questions, while dimension 5 and above are again easier.

Two-dimensional manifolds. This is a classical area presented in many textbooks. There is a well known complete classification of compact 2-dimensional manifolds, also known as 2-dimensional closed surfaces (see [**FW99**] for a nice pictorial proof of the classification, which assumes that the manifold is triangulated, though).

We assume that the considered 2-dimensional manifold is connected (if not, consider each component separately). Up to homeomorphism, such manifolds fall into two groups: the orientable ones, consisting of the sphere S^2, the torus, the double torus, etc. (all are obtained by attaching handles to the sphere), and the unorientable ones, which are the projective plane, the Klein bottle, and others that can be constructed by attaching handles to one of these two. The compact 2-manifolds with boundary are obtained from the ones without boundary by cutting out finitely many disjoint disks.

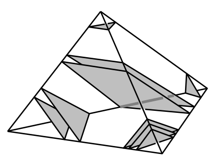
Three-dimensional manifolds. Lots of knowledge has been accumulated about these, although outstanding problems still remain, many of them of algorithmic nature. The techniques are mostly very specific to this area, which is almost separate from the rest of topology.

A basic tool for three-dimensional manifolds is the *theory of normal surfaces*, which was originally developed by Haken in order to get

an algorithm for testing whether a given (piecewise linear) embedding $S^1 \to S^3$ is knotted.

Let us make a digression and sketch the main ideas of this beautiful algorithm. It is well known that an embedding of S^1 in S^3 is unknotted exactly if there is a disk D embedded in S^3 such that the embedded S^1 is its boundary. Given an embedded S^1, the algorithm considers the corresponding knot manifold M (i.e., S^3 minus the slightly thickened knot) and searches for a suitable embedded disk D in M (such that the boundary of D is contained in the torus bounding M and goes "once around" it). We assume that M is triangulated, and that the triangulation has t tetrahedra.

Assuming that such a D exist, we pull it taut within M and then perturb it slightly. Then the intersection of D with each tetrahedron is of the type shown in the picture:



Each component of the intersection is either a triangle, one of four possible types, or a quadrilateral, one of three possible types (where two types of quadrilaterals cannot coexist in a single tetrahedron). (A *normal surface* in general is a surface in M whose intersections with all tetrahedra have the just explained form.)

For every tetrahedron τ, we write down 7 nonnegative integers, corresponding to the number of components of the intersection $D \cap \tau$ of each of the 7 possible types, where the types are enumerated in some fixed order. Doing this for every tetrahedron, we obtain a vector $v_D \in \mathbb{Z}^{7t}_{\geq 0}$, the *coordinate vector* of D.

The coordinate vector describes D uniquely up to a continuous deformation within M. Moreover, given a vector $v \in \mathbb{Z}^{7t}_{\geq 0}$, we can check whether it actually describes the desired disk D witnessing unknotedness.

Of course, this does not yet give an algorithm, since there are infinitely many nonnegative integer vectors. We observe that, by far,

not all such vectors actually describe a surface in M; one necessary condition is that the vectors be compatible on the boundaries of the tetrahedra. That is, if T is a triangle bounding a tetrahedron τ, then the components of the coordinate vector v_D corresponding to τ determine the number and type of segments of $D \cap T$. Since each triangle, except for those on the boundary of M, bounds two tetrahedra, it imposes a condition on v_D, which can be expressed as several *homogeneous* linear equations for the components of v_D.

Thus, it suffices to look for the desired coordinate vector v_D only among the (nonnegative) solutions of the linear system expressing the compatibility conditions.

Now, crucially, it can be shown (nontrivially) that we can restrict the search to *fundamental solutions*, where a solution v is fundamental if it cannot be written as $v = v_1 + v_2$, where both v_1 and v_2 are nonzero solutions.

A basic result in the theory of integral cones asserts that a system of m linear equations (and possibly inequalities) with n unknowns and with integer coefficients has only a finite number of fundamental nonnegative solutions, and this number can be bounded by a function of m, n, and of the size of the coefficients. The fundamental solutions can also be enumerated algorithmically (see, e.g., [**Zie94**]).

The algorithm thus consists in generating all fundamental solutions and testing each of them if it provides the desired disk D. We refer to Hass et al. [**HLP99**] for a more detailed presentation.

Algorithms obtained by the method of normal surfaces are typically at least exponential, and the existence of polynomial-time algorithms for some of the problems (e.g., recognizing whether a given simplicial complex is a triangulation of S^3, or detecting knotedness) presents fascinating questions.

Another, more recent tool for studying 3-mainfolds, comes from *Thurston's geometrization conjecture*. Roughly speaking, the conjecture asserts that every compact orientable 3-manifold (without boundary) can be cut into finitely many pieces so that each of the pieces can be endowed with one of 8 very special geometries.

Here "geometry" basically means metric; more precisely, a Riemannian metric. To give a simple example, let us consider 2-dimensional manifolds first. The sphere S^2 with the geodesic metric (measuring shortest distance along the surface) has constant positive curvature. It turns out that S^2 is the *only* orientable 2-manifold that can be given a geometry of constant positive curvature. The torus, for example, admits a flat (zero-curvature) Euclidean geometry, and the double torus a hyperbolic, constant negative curvature geometry. The *uniformization theorem* implies that every compact 2-manifold can be endowed with one of these three types of geometry.

The geometrization conjecture is a similar kind of statement, except that the 3-manifold must in general be cut into "geometrizable" pieces, and that there are 8 possible geometries, some of them quite exotic-looking. But knowing the geometry provides very good understanding of the considered manifold.

The geometrization conjecture was proved by Perel'man in 2003, building on the work of Thurston and Hamilton, by an approach through a nonlinear partial differential equation (Perel'man's techniques have also had great impact in PDE's).

The most celebrated result of Perel'man, which is a consequence of the geometrization conjecture but does not need the full strength of it, was a proof of the **Poincaré conjecture**.

Poincaré was initially wondering whether the homology groups characterize the 3-sphere among compact 3-manifolds. Soon he found a counterexample, known as the *Poincaré homology sphere*, one of the important and elegant examples in topology, which has a nontrivial fundamental group and thus cannot be homeomorphic to S^3. So next he asked whether every compact simply connected 3-manifold is homeomorphic to S^3, and this is what became known as the Poincaré conjecture (and one of the Clay Institute's "Millennium Problems").

A (nontrivial) reformulation of the Poincaré conjecture is whether every 3-manifold homotopy equivalent to S^3 must also be homeomorphic to S^3, and in this form, it makes sense for every dimension n. This is the *generalized Poincaré conjecture*.

Interestingly, the $n = 3$ was the last case to be solved: the generalized Poincaré conjecture for $n \geq 5$ was proved in the 1960s, and for $n = 4$ in 1982.

Dimensions 5 and more. Five- and higher-dimensional manifolds were originally thought to be even more difficult than four-dimensional ones, but after a breakthrough of Smale in the 1960s, they are now much better understood, especially in the differentiable case. Smale proved a celebrated result known as the *h-cobordism* theorem, claiming that certain kind of equivalence of simply connected manifolds in dimension at least 5 actually implies homeomorphism. An immediate consequence was the generalized Poincaré conjecture for $n \geq 5$.

Later on, a *surgery theory* was developed, which provides an algebraic classification of in these dimensions manifolds. Since homeomorphism of two given triangulated manifolds of every dimension $n \geq 4$ is algorithmically undecidable, the classification is "inefficient" in a sense, but it has been used with success to solve various concrete problems. For example, these methods have been used in the classification of exotic spheres (nonequivalent differential structures in S^n), as well as in results related to the Hauptvermutung and triangulability of manifolds mentioned earlier.

Four-dimensional manifolds. These are the most problematic ones. Some of the higher-dimensional theory was extended, mainly by Freedman, to work with 4-dimensional *topological* manifolds (and in this way the generalized Poincaré conjecture was also proved for $n = 4$), but not at all for differentiable manifolds, and basic questions about them remain unresolved.

Scorpan's book [**Sco05**] is devoted to 4-manifolds, has many nice pictures, and in the first part it also explains material around the h-cobordism theorem, how things work for dimension 5 and more and why they fail in dimension 4.

15. Literature

A usual source for general topology is Kelley [**Kel75**]; our favorite book is Engelking [**Eng89**]. For an accessible introduction to combinatorial and algebraic topology the best recommendation we can provide are two books of Prasolov [**Pra06, Pra07**] (in [**Pra95**] the same author explains quite sophisticated topological examples almost entirely by intuitive pictorial problems and challenges). A standard, on-line accessible, and quite readable (not to be confused with easy) textbook of algebraic topology is Hatcher [**Hat01**].

For manifolds, differential topology and such the literature is vast, and some references have already been given earlier. For the beginning it is advisable to read Milnor; all of his books and lecture notes, although quite old by now, seem hard to beat in quality and accessibility.

A good, if a bit dated, introduction to knots and 3-dimensional topology is Rolfsen [**Rol90**]. Hempel [**Hem76**] is a book on 3-manifolds. Category theory is presented in [**ML98**] by its inventor; another more recent book is Adámek et al. [**AHS06**].

Acknowledgment. This chapter was inspired, among others, by lecture notes of Günter M. Ziegler (Topologie, in German). We would like to thank Martin Čadek for expert advice concerning the planned contents, and Martin Tancer and, especially, Günter M. Ziegler for careful reading and useful comments.

Bibliography

[AH00] R. Aharoni and P. Haxell. Hall's theorem for hypergraphs. *J. Graph Theory*, 35(2):83–88, 2000.

[Aha01] R. Aharoni. Ryser's conjecture for tripartite 3-graphs. *Combinatorica*, 21(1):1–4, 2001.

[AHS06] J. Adámek, H. Herrlich, and G. E. Strecker. Abstract and concrete categories: the joy of cats. *Repr. Theory Appl. Categ.*, 17, 2006. Reprint of the 1990 original [Wiley, New York], also available on-line.

[Ber03] M. Berger. *A Panoramic View of Riemannian Geometry*. Springer-Verlag, Berlin, 2003.

[Bjö95] A. Björner. Topological methods. In R. Graham, M. Grötschel, and L. Lovász, editors, *Handbook of Combinatorics*, volume II, chapter 34, pages 1819–1872. North-Holland, Amsterdam, 1995.

[Bjö03] A. Björner. Nerves, fibers and homotopy groups. *J. Combin. Theory Ser. A*, 103:88–93, 2003.

[CdVGG12] É. Colin de Verdière, G. Ginot, and X. Goaoc. Multinerves and Helly numbers of acyclic families. In *Proc. ACM Sympos. Comput. Geometry (SoCG'12)*, pages 209–217. ACM, New York, 2012.

[ČKM+14] M. Čadek, M. Krčál, J. Matoušek, F. Sergeraert, L. Vokřínek, and U. Wagner. Computing all maps into a sphere. *J. ACM*, 61(3), 2014. Article No. 17. Preprint in arXiv:1105.6257.

[DLRS10] J. A. De Loera, J. Rambau, and F. Santos. *Triangulations*, volume 25 of *Algorithms and Computation in Mathematics*. Springer-Verlag, Berlin, 2010.

[Eng89] R. Engelking. *General Topology*. Heldermann Verlag, Berlin, second edition, 1989.

[Fri12] G. Friedman. An elementary illustrated introduction to simplicial sets. *Rocky Mountain J. Math.*, 42(2):353–423, 2012.

[FW99] G. K. Francis and J. R. Weeks. Conway's ZIP proof. *Amer. Math. Monthly*, 106(5):393–399, 1999.

[Hat01] A. Hatcher. *Algebraic Topology*. Cambridge University Press, Cambridge, 2001. Electronic version available at `http://math.cornell.edu/~hatcher#AT1`.

[Hem76] J. Hempel. 3-*Manifolds*. Princeton University Press, Princeton, NJ, 1976. Ann. of Math. Studies, No. 86.

[HKR13] M. Herlihy, D. Kozlov, and S. Rajsbaum. *Distributed Computing through Combinatorial Topology*. Morgan Kaufmann, 2013.

[HLP99] J. Hass, J. C. Lagarias, and N. Pippenger. The computational complexity of knot and link problems. *J. ACM*, 46(2):185–211, 1999.

[Kal12] G. Kalai. Happy birthday Ron Aharoni! Blog, `https://gilkalai.wordpress.com/2012/11/25/happy-birthday-ron-aharoni/`, 2012.

[Kel75] J. L. Kelley. *General Topology*. Springer-Verlag, New York–Berlin, 1975.

[Lut15] F. Lutz. *Triangulated Manifolds*. Springer, Berlin, 2015. In production. Some of the chapters, by several contributors, available at `http://page.math.tu-berlin.de/~lutz/`.

[Lyo03] D. W. Lyons. An elementary introduction to the Hopf fibration. *Math. Mag.*, 76(2):87–98, 2003.

[Mat03] J. Matoušek. *Using the Borsuk–Ulam Theorem*. Springer, Berlin, 2003.

[Mes01] R. Meshulam. The clique complex and hypergraph matching. *Combinatorica*, 21(1):89–94, 2001.

[ML98] S. Mac Lane. *Categories for the Working Mathematician*. Springer-Verlag, New York, second edition, 1998.

[Mun84] J. R. Munkres. *Elements of Algebraic Topology*. Addison-Wesley, Reading, MA, 1984.

[Pra95] V. V. Prasolov. *Intuitive Topology*. American Mathematical Society, Providence, RI, 1995.

[Pra06] V. V. Prasolov. *Elements of Combinatorial and Differential Topology*. American Mathematical Society, Providence, RI, 2006.

[Pra07] V. V. Prasolov. *Elements of Homology Theory*. American Mathematical Society, Providence, RI, 2007.

[Rol90] D. Rolfsen. *Knots and Links*. Publish or Perish, Inc., Houston, TX, 1990. Corrected reprint of the 1976 original.

[RS12] J. Rubio and F. Sergeraert. Constructive homological algebra and applications. Preprint, arXiv:1208.3816, 2012. Written in 2006 for a MAP Summer School at the University of Genova.

[Rud01] Yu. B. Rudyak. Piecewise linear structures on topological manifolds. Preprint arXiv:math/0105047, 2001.

[Sco05] A. Scorpan. *The Wild World of 4-Manifolds*. American Mathematical Society, Providence, RI, 2005.

[Wac07] M. L. Wachs. Poset topology: tools and applications. In *Geometric Combinatorics*, pages 497–615. American Mathematical Society, Providence, RI, 2007.

[Zie94] G. M. Ziegler. *Lectures on Polytopes*, volume 152 of *Graduate Texts in Mathematics*. Springer-Verlag, Heidelberg, 1994. Corrected and revised printing 1998.

Index

$\int_D f \, \mathrm{d}\mu$, 25
$\|x\|$ (Euclidean norm), 41
A_t (t-neighborhood), 47
$\langle u, v \rangle$ (inner product), 87
$\|u\|$ (norm), 87
$V \oplus W$ (direct sum), 146
$V \otimes W$ (tensor product), 149
$R[x_1, \ldots, x_n]$ (polynomial ring), 174
$\langle \mathcal{F} \rangle$ (ideal generated by $\mathcal{F}$), 186
$\sqrt{I}$ (radical), 190
$\Bbbk[X]$ (coordinate ring), 197
$X \cong Y$ (homeomophic), 237
∂ (boundary operator), 303
∂Y (boundary), 239
$f \sim g$ (homotopic maps), 254
$[X, Y]$ (homotopy classes), 255
$X \simeq Y$ (homotopy equivalence), 256
X/A (quotient space), 262
$X * Y$ (join), 264
$|K|$ (polyhedron), 275
$[f]_*$ (pointed homotopy class), 291
χ-distribution, 71
$\chi_a(x)$ (character), 92
$\chi(G)$ (chromatic number), 260
$\gamma(A)$ (Gaussian measure), 54
$\lambda(A)$ (Lebesgue measure), 10
$\lambda^*(A)$ (outer Lebesgue measure), 6
$\pi_1(X)$ (fundamental group), 292
$\pi_k(X)$ (k-th homotopy group), 297
σ-algebra, 12
σ-field, 12
σ-finite measure, 30

abstract nonsense, 271
additivity
  countable, 8
  finite, 8
affine variety, 185
Alexander duality, 321
algebra, group, 144
algebraic closure, 188
algebraic geometry, 186
algebraic topology, 253
algebraically closed field, 186
algebraically independent numbers, 181
algorithm
  division, 212
  Haken's, 326
almost everywhere, 14
alternating representation, 144
annulus, 254
antipodal map, 258
approximate identity, 134
atlas, 323
axiom of choice, 13, 125, 250
axioms, Kolmogorov's, 33

$B_k(K;\mathbb{Z}_2)$ (k-boundaries), 304
B^n (Euclidean unit ball), 236
$B(x,r)$ (ball), 43
ball, 236
  volume, 42, 60
Banach–Tarski paradox, 9
barycentric subdivision, 311
base, 238
basepoint, 291
basis
  Fourier, 93
  Gröbner, 213
Behrend's construction, 105
Benford's law, 35
Bernstein method, 74
Betti number, 307
Bézout's inequality, 197
  generalized, 208
  higher-dimensional, 219
Bing's house, 257
body, convex, 41
Boolean function, 99, 113
  monotone, 117
Borel set, 12
Borsuk–Ulam theorem, 258
bottle, Klein, 284
boundary, 239
  k-boundary, 304
boundary operator, 303
Brouwer's fixed-point theorem, 318
Brunn's inequality, 44

$C_k(K;\mathbb{Z}_2)$ (k-chains), 302
Cantor set, 10, 241
cap, spherical, 64
categorical limit, 269
categorical product, 269
category, 267
  opposite, 270
Cauchy equation, 125
cell complex, 280
central limit theorem, 55
chain
  k-chain, 303
chain complex, 306
chain map, 306
character, 88, 124, 150, 152
  trivial, 89
Chebyshev's inequality, 75
Cheeger inequality, 94
Chernoff's inequality, 77
Chevalley's theorem, 205
choice, axiom of, 13, 125, 250
chromatic number, 260
circle group, 87
$\mathrm{cl}\, Y$ (closure), 239
class, homology, 304
class function, 151
clique complex, 278
closed set, 239
closure, 239
  algebraic, 188
coboundary operator, 308
code, locally decodable, 169
coefficient, 174
  leading, 187
coefficients, Fourier, 128
cohomology group, 308
communication complexity, 163
compact set, 247
compact space, 247
complement, orthogonal, 110, 147
complete measure, 30
completion, projective, 216
complex
  cell, 280
  chain, 306
  clique, 278
  CW, 280
  flag, 278
  independence, 279
  order, 279
  simplicial, 271
complexity, communication, 163
component
  connected, 241
  irreducible, 201
concave function, 41
cone, 265
conjecture
  geometrization, 327
  log-rank, 164
  Poincaré, 328
conjugate elements, 151
connected space, 240
connected component, 241

constant-degree expander, 63
constructible set, 205
construction, Behrend's, 105
contagious vanishing, 185
content
  Jordan, 2
  Minkowski, 48
continuous map, 239
contractible space, 257
contravariant functor, 309
convergence in distribution, 55
convex body, 41
convex function, 41
convex hull, 41
convex polyhedron, 41
convex polytope, 41
  enclosing S^n, 65
convex set, 41
convolution, 106, 158
convolution as multiplication, 107
conv X (convex hull), 41
coordinate ring, 197
coordinates, homogeneous, 215
coproduct, 270
correspondence
  ideal–variety, 190
countable additivity, 8
counting measure, 16
covariant functor, 309
cover, open, 247
cube
  Hamming, 63
  Hilbert, 246
cup product, 309
curve, moment, 275
CW complex, 280
cycle
  k-cycle, 304

Daniell integral, 5
deformation retract, 256
deg f (degree of a polynomial), 174
degree
  of a polynomial, 174
  of a variety, 208
deleted product, 286
delta function, Kronecker, 94
ε-dense set, 64
dense set, 244
density, 14
derivative
  Hasse, 182
  partial, 182
determinant, Jacobian, 222
determinantal variety, 204
diagram, Ferrer's, 161
diameter, 17
dictatorship, 113
differentiable manifold, 323
dimension
  Hausdorff, 18
  of a simplicial complex, 272
  of a variety, 207
  of a representation, 143
Dirac measure, 16
direct sum, 146
Dirichlet's theorem, 131
Dirichlet kernel, 131
discrete topology, 242
disjoint union, 266
distribution
  Gaussian, 53
  normal, 54
    tail estimates, 57
  probability, 34
  stable, 60
  standard normal, 53
  subgaussian, 76
  uniform, 32
division algorithm, 212
domain
  frequency, 128
  invariance of, 318
  time, 128
domain, integral, 200
dominated convergence theorem, 29
duality
  Alexander, 321
  finite Pontryagin, 89
  Pontryagin, 126
dunce hat, 263

$e(x)$, 88
$\mathbb{E}[X]$ (expectation), 35
elementary event, 32
elements, conjugate, 151

embedding, 283
  Segre, 217
entropy, 123
epimorphism, 268
equation, Cauchy, 125
equidecomposable sets, 9
equivalence, homotopy, 256
equivalent representations, 143
event, 32
  elementary, 32
everywhere, almost, 14
exhaustion, method of, 2
exotic sphere, 324
expander, constant-degree, 63
expectation, 35, 89
extended real numbers, 6

$\mathbb{F}_q$ (finite field), 173
face, 273
factorization, unique, 200
fast multiplication of polynomials, 108
fast Fourier transform, 86, 98
Fatou's lemma, 27
Fejér kernel, 134
Fejér's theorem, 133
Ferrer's diagram, 161
field, 173
  algebraically closed, 186
finite additivity, 8
finitely presented group, 294
flag complex, 278
$f_{\rm low}$ (low-degree part), 115
Fourier inversion, noncommutative, 158
Fourier basis, 93
Fourier coefficients, 128
Fourier series, 85, 129, 130
Fourier transform, 95, 127, 150
  fast, 86, 98
  inverse, 96, 129
  noncommutative, 151
free probability, 36
frequencies, 86
frequency domain, 128
Fubini's theorem, 31
function
  Boolean, 99, 113
    monotone, 117
  class, 151
  concave, 41
  convex, 41
  Hilbert, 198
  Lipschitz, 66
  log-concave, 53
  majority, 113
  measurable, 22, 23
  simple, 23
  unimodal, 45
functional, Laplace, 78
functor, 293
  contravariant, 309
  covariant, 309
fundamental group, 292

G-linear map, 143
Gauss map, 286
Gauss sum, 112
Gaussian distribution, 53
Gaussian measure, 54
  spherical shell, 70, 77
Gaussian measure concentration, 68
general linear group, 142
general topology, 242
geometric realization, 274
geometric simplicial complex, 274
geometrization conjecture, 327
geometry, algebraic, 186
$\mathrm{GL}(n, \mathbb{K})$, 20
$\mathrm{GL}(V)$, 142
graded lexicographic ordering, 209
gradient, 182, 221
graph
  k-colorable, 250
  Kneser, 259
Gröbner basis, 213
Gromov's sphere waist theorem, 67
group
  circle, 87
  cohomology, 308
  finitely presented, 294
  fundamental, 292
  general linear, 142
  homology, 301
  homotopy, 299

  LCA, 126
  matrix, 20
  representation, 142
  symmetric, 144
    representations, 160
  topological, 21
group algebra, 144

$H_k(X; R)$ (homology group), 301
Haar's theorem, 21
Haar measure, 19, 127
Haefliger–Weber theorem, 287
Haken's algorithm, 326
Hamming cube, 63
Hamming metric, 63
Hasse derivative, 182
hat, dunce, 263
Hauptvermutung, 315
Hausdorff dimension, 18
Hausdorff measure, 17
Hausdorff space, 244
Heine–Borel theorem, 7, 251
Hessian matrix, 228
$\mathrm{HF}_R(d)$ (Hilbert function), 198
Hilbert's third problem, 1
Hilbert basis theorem, 187
Hilbert cube, 246
Hilbert function, 198
Hilbert polynomial, 208
$\mathrm{Hom}(X, Y)$ (morphisms), 268
homeomorphism, 237
  and null set, 15
homogeneous coordinates, 215
homogeneous ideal, 216
homogeneous polynomial, 216
homology class, 304
homology group, 301
homotopic maps, 254
homotopy, 254
  pointed, 291
homotopy equivalence, 256
homotopy group, 299
  of a sphere, 299
homotopy type, 256
Hopf map, 300
house, Bing's, 257
hull, convex, 41
Hurewicz's theorem, 320
hypercontractive inequality, 119
hypercontractive operator, 122

I_n (identity matrix), 152
$I(S)$ (vanishing ideal), 190
ideal, 186
  homogeneous, 216
  monomial, 209
  prime, 201
  radical, 190
ideal–variety correspondence, 190
identity, approximate, 134
independence complex, 279
induced subcomplex, 272
inequality
  Bézout's, 197
  Bézout's
    generalized, 208
    higher-dimensional, 219
  Brunn's, 44
  Chebyshev's, 75
  Cheeger, 94
  Chernoff's, 77
  hypercontractive, 119
  isoperimetric, 48
  Jensen's, 79
  log-Sobolev, 123
  Markov's, 75
  Prékopa–Leindler, 50
  Young's, 109
influence, 113
  total, 114
$\mathrm{Inf}_k(f)$ (influence), 113
inner product, 87, 90
$\mathrm{int}\, Y$ (interior), 239
integral
  Daniell, 5
  Lebesgue, 3, 25
  Riemann, 2
integral domain, 200
interior, 239
invariance of domain, 318
invariant subspace, 145
inverse Fourier transform, 96, 129
inversion
  Fourier, noncommutative, 158
irreducible representation, 145
irreducible component, 201

irreducible polynomial, 188
irreducible variety, 201, 203
isomorphism, 203, 268
  of simplicial complexes, 273
isoperimetric inequality, 48
isoperimetric problem, 46, 63

Jacobian determinant, 222
Jensen's inequality, 79
Johnson–Lindenstrauss lemma, 72
join, 264, 278
joint, 181
Jordan curve theorem, 321
Jordan content, 2

kernel
  Dirichlet, 131
  Fejér, 134
KKL theorem, 113
Klein bottle, 284
Kneser graph, 259
Kolmogorov's axioms, 33
Kronecker delta function, 94
Kuratowski's theorem, 283

$L^1(G)$ (integrable functions), 127
ℓ_1-product, 78
Lévy's lemma, 66
Laplace functional, 78
law, Benford's, 35
LCA group, 126
leading coefficient, 187
leading monomial, 209
Lebesgue covering lemma, 314
Lebesgue density theorem, 15
Lebesgue integral, 3, 25
Lebesgue measure, 3, 10
Legendre symbol, 112
lemma
  Fatou's, 27
  Johnson–Lindenstrauss, 72
  Lévy's, 66
  Lebesgue covering, 314
  random projection, 73
  Riemann–Lebesgue, 135
limit, categorical, 269
line, Sorgenfrey, 243
Lipschitz function, 66
Littlewood's principles, 11, 23

locally compact space, 21
locally decodable code, 169
log-concave function, 53
log-rank conjecture, 164
log-Sobolev inequality, 123
long ray, 243
longest increasing subsequence, 168
Lovász–Kneser theorem, 260
Lyusternik–Schnirel'man theorem, 258

majority function, 113
manifold, 321
  differentiable, 323
  smooth, 323
  topological, 322
  with boundary, 322
map
  antipodal, 258
  chain, 306
  continuous, 239
  G-linear, 143
  Gauss, 286
  Hopf, 300
  nullhomotopic, 255
  of pairs, 292
  open, 240
  pointed, 291
  rational, 206
  regular, 203
  simplicial, 273
maps, homotopic, 254
Markov's inequality, 75
Maschke's theorem, 146
matching, perfect, 176
matrix
  Hessian, 228
  Sylvester, 192
  unitary, 148
matrix group, 20
measurable function, 22, 23
measurable rectangle, 30
measurable set, 10
measure, 16
  complete, 30
  counting, 16
  Dirac, 16
  Gaussian, 54

Haar, 19, 127
Hausdorff, 17
Lebesgue, 3, 10
outer Lebesgue, 6
probability, 34
product, 30
σ-finite, 30
measure concentration
Gaussian, 68
Hamming cube, 77
product space, 78
sphere, 62, 69
various spaces, 62
measure space, 5, 16
median, 66
method
Bernstein, 74
of exhaustion, 2
metric, 236
Hamming, 63
metric space, 236
metrizable space, 242
Minkowski content, 48
Minkowski sum, 45
model, vector, 40
module, Specht, 161
moment curve, 275
monomial, 174
leading, 209
number of, 180
monomial ideal, 209
monomial ordering, 209
monomorphism, 268
monotone Boolean function, 117
monotone convergence theorem, 28
morphism, 267
of varieties, 202, 217

$N(\mu, \sigma^2)$ (normal distribution), 60
negative part, 23
neighborhood, 238
t-neighborhood, 47
nerve theorem, 279
Noetherian ring, 187
noise operator, 122
non-measurable set, 13
noncommutative Fourier transform, 151
nonsense, abstract, 271
nonsingular zero, 222
norm, 87, 90
normal distribution, 54
tail estimates, 57
normal space, 244
normal surface, 325
null set, 14
nullhomotopic map, 255
Nullstellensatz, 188
number
Betti, 307
chromatic, 260
of components of a variety, 226
of monomials, 180
numbers, algebraically independent, 181

open cover, 247
open map, 240
open set, 237
operator
boundary, 303
coboundary, 308
hypercontractive, 122
noise, 122
opposite category, 270
order complex, 279
ordering
graded lexicographic, 209
monomial, 209
orthogonal complement, 110, 147
outer Lebesgue measure, 6

paracompact space, 252
paradox, Banach–Tarski, 9
Parseval's theorem, 96
part
negative, 23
positive, 23
torsion, 307
partial derivative, 182
partitions of n, 161
path, 241
path-connected space, 241
perfect matching, 176
permutation representation, 144
Plancherel's theorem, 96

plane
  projective, 284
  Sorgenfrey, 243
Poincaré conjecture, 328
point, random, 31
point-set topology, 242
pointed homotopy, 291
pointed map, 291
pointed space, 291
Poisson summation formula, 111, 136
Polish space, 245
polyhedron
  convex, 41
  of a simplicial complex, 275
polynomial
  degree, 174
  fast multiplication, 108
  Hilbert, 208
  homogeneous, 216
  irreducible, 188
  zero, 174
polynomial identity testing, 179
polytope, convex, 41
  enclosing S^n, 65
Pontryagin dual, 89
Pontryagin duality theorem, 126
positive part, 23
Prékopa–Leindler inequality, 50
prime ideal, 201
principle
  uncertainty, 97
principles, Littlewood's, 11, 23
probability, 33
  free, 36
probability distribution, 34
probability measure, 34
probability space, 32
problem
  Hilbert's third, 1
  isoperimetric, 46, 63
  undecidable, 253, 294
product
  categorical, 269
  cup, 309
  deleted, 286
  inner, 87
  ℓ_1, 78
  tensor, 149
product space, measure concentration, 78
product, inner, 90
product measure, 30
product topology, 249, 265
projection, 207
  random, 73
projection theorem, 218
projective completion, 216
projective plane, 284
projective space, 214
property testing, 99

quadratic residue, 111
quotient space, 262

$\overline{\mathbb{R}}$ (extended reals), 6
Rabinowitsch trick, 190
radical ideal, 190
random projection lemma, 73
random point, 31
  on S^n, 41, 59
random projection, 73
random variable, 35
rational map, 206
ray, long, 243
real numbers, extended, 6
realization, geometric, 274
rectangle, measurable, 30
regular map, 203
regular representation, 144
regular space, 244
rejection sampling, 42
representation, 142
  alternating, 144
  dimension, 143
  irreducible, 145
  permutation, 144
  regular, 144
  symmetric group, 160
  trivial, 144
representations, equivalent, 143
$\mathrm{Res}(f, g, x)$ (resultant), 192
residue, quadratic, 111
resultant, 192
retract, deformation, 256
retraction, 319

Riemann integral, 2
Riemann–Lebesgue lemma, 135
ring, 173
  coordinate, 197
  Noetherian, 187
Roth's theorem, 101

S_n (symmetric group), 144
S^n (Euclidean unit sphere), 236
sample space, 32
sampling, rejection, 42
Schur's theorem, 153
Schwartz–Zippel theorem, 175
second-countable space, 245
Segre embedding, 217
semialgebraic set, 186
separable space, 245
series, Fourier, 85, 129, 130
set
  Borel, 12
  Cantor, 10, 241
  closed, 239
  compact, 247
  constructible, 205
  convex, 41
  dense, 244
  ε-dense, 64
  measurable, 10
  non-measurable, 13
  null, 14
  open, 237
  semialgebraic, 186
  simplicial, 281
  Smith–Volterra–Cantor, 15
  Vitali, 9
sets, equidecomposable, 9
sign pattern, 231
simple function, 23
simplex, 273
simplicial approximation theorem, 312
simplicial complex, 271
  geometric, 274
  isomorphism, 273
  polyhedron, 275
simplicial map, 273
simplicial set, 281
simply connected space, 293
Smith–Volterra–Cantor set, 15
smooth manifold, 323
SO($n, \mathbb{R}$), 20
Sorgenfrey line, 243
Sorgenfrey plane, 243
space
  compact, 247
  connected, 240
  contractible, 257
  Hausdorff, 244
  locally compact, 21
  measure, 5, 16
  metric, 236
  metrizable, 242
  normal, 244
  paracompact, 252
  path-connected, 241
  pointed, 291
  Polish, 245
  probability, 32
  projective, 214
  quotient, 262
  regular, 244
  sample, 32
  second-countable, 245
  separable, 245
  simply connected, 293
  T_i, 243
  topological, 237
Specht module, 161
sphere, 236
  exotic, 324
  homotopy groups, 299
  waist theorem, 67
spherical shell, Gaussian measure, 70, 77
spherical cap, 64
stable distribution, 60
standard normal distribution, 53
standard tableau, 162
subadditivity, 8
subbase, 239
subcomplex, 272
  induced, 272
subcover, 247
subdivision, 310
  barycentric, 311
subgaussian distribution, 76

subrepresentation, 145
subsequence, longest increasing, 168
subspace, 238
  invariant, 145
sum
  direct, 146
  Gauss, 112
  Minkowski, 45
summation formula, Poisson, 111, 136
supp(f) (support), 98
support, 98, 276
surface
  normal, 325
  two-dimensional, classification, 325
suspension, 265
Sylvester matrix, 192
symbol, Legendre, 112
symmetric group, 144
  representations, 160

$\mathbb{T}$ (circle group), 87
T_i space, 243
tableau, 161
  standard, 162
tabloid, 161
tensor power trick, 123
tensor product, 149
term, 174
testing
  property, 99
testing, polynomial identity, 179
theorem
  Borsuk–Ulam, 258
  Brouwer's fixed-point, 318
  central limit, 55
  Chevalley's, 205
  Dirichlet's, 131
  dominated convergence, 29
  Fejér's, 133
  Fubini's, 31
  Gromov's sphere waist, 67
  Haar's, 21
  Haefliger–Weber, 287
  Heine–Borel, 7, 251
  Hilbert basis, 187
  Hurewicz's, 320
  Jordan curve, 321
  KKL, 113
  Kuratowski's, 283
  Lebesgue density, 15
  Lovász–Kneser, 260
  Lyusternik–Schnirel'man, 258
  Maschke's, 146
  monotone convergence, 28
  nerve, 279
  Parseval's, 96
  Plancherel's, 96
  Pontryagin duality, 126
  projection, 218
  Roth's, 101
  Schur's, 153
  Schwartz–Zippel, 175
  simplicial approximation, 312
  Tietze extension, 245
  Tonelli's, 31
  Tychonoff's, 250
  Urysohn metrization, 245
  van Kampen–Flores, 285
Tietze extension theorem, 245
time domain, 128
Tonelli's theorem, 31
topological group, 21
topological manifold, 322
topological space, 237
topology, 237
  algebraic, 253
  discrete, 242
  general, 242
  point-set, 242
  product, 249, 265
  Zariski, 202, 243
torsion part, 307
torus, 263
total influence, 114
Tr A (trace), 151
trace, 151
transform
  Fourier, 95, 127, 150
triangulation, 277
trick
  Rabinowitsch, 190
  tensor power, 123
trivial character, 89

trivial representation, 144
Tychonoff's theorem, 250
type, homotopy, 256

uncertainty principle, 97
undecidable problem, 253, 294
uniform distribution, 32
unimodal function, 45
union, disjoint, 266
unique factorization, 200
unitarity, Weyl's, 148
unitary matrix, 148
Urysohn metrization theorem, 245

$V(\mathcal{F})$ (variety of $\mathcal{F}$), 185
$V(K)$ (vertex set), 271
van Kampen–Flores theorem, 285
vanishing, contagious, 185
variable, random, 35
variety
  affine, 185
  degree, 208
  determinantal, 204
  dimension, 207
  irreducible, 201, 203
  morphism, 217
  number of compoments, 226
vector model, 40
vertex, 271
Vitali set, 9
volume of a ball, 42, 60

wavelets, 86
wedge, 292
Weyl's unitarity, 148

Young's inequality, 109

$\mathbb{Z}_2$-map, 286
$\mathbb{Z}_2$-space, 286
$Z_k(K;\mathbb{Z}_2)$ (k-cycles), 304
Zariski topology, 202, 243
zero, nonsingular, 222
zero polynomial, 174

Selected Published Titles in This Series

75 **Ida Kantor, Jiří Matoušek, and Robert Šámal,** Mathematics++, 2015
73 **Bruce M. Landman and Aaron Robertson,** Ramsey Theory on the Integers, Second Edition, 2014
72 **Mark Kot,** A First Course in the Calculus of Variations, 2014
71 **Joel Spencer,** Asymptopia, 2014
70 **Lasse Rempe-Gillen and Rebecca Waldecker,** Primality Testing for Beginners, 2014
69 **Mark Levi,** Classical Mechanics with Calculus of Variations and Optimal Control, 2014
68 **Samuel S. Wagstaff, Jr.,** The Joy of Factoring, 2013
67 **Emily H. Moore and Harriet S. Pollatsek,** Difference Sets, 2013
66 **Thomas Garrity, Richard Belshoff, Lynette Boos, Ryan Brown, Carl Lienert, David Murphy, Junalyn Navarra-Madsen, Pedro Poitevin, Shawn Robinson, Brian Snyder, and Caryn Werner,** Algebraic Geometry, 2013
65 **Victor H. Moll,** Numbers and Functions, 2012
64 **A. B. Sossinsky,** Geometries, 2012
63 **María Cristina Pereyra and Lesley A. Ward,** Harmonic Analysis, 2012
62 **Rebecca Weber,** Computability Theory, 2012
61 **Anthony Bonato and Richard J. Nowakowski,** The Game of Cops and Robbers on Graphs, 2011
60 **Richard Evan Schwartz,** Mostly Surfaces, 2011
59 **Pavel Etingof, Oleg Golberg, Sebastian Hensel, Tiankai Liu, Alex Schwendner, Dmitry Vaintrob, and Elena Yudovina,** Introduction to Representation Theory, 2011
58 **Álvaro Lozano-Robledo,** Elliptic Curves, Modular Forms, and Their L-functions, 2011
57 **Charles M. Grinstead, William P. Peterson, and J. Laurie Snell,** Probability Tales, 2011
56 **Julia Garibaldi, Alex Iosevich, and Steven Senger,** The Erdős Distance Problem, 2011
55 **Gregory F. Lawler,** Random Walk and the Heat Equation, 2010
54 **Alex Kasman,** Glimpses of Soliton Theory, 2010
53 **Jiří Matoušek,** Thirty-three Miniatures, 2010

For a complete list of titles in this series, visit the
AMS Bookstore at **www.ams.org/bookstore/stmlseries/**.





Photo courtesy of ETH Zürich, Giulia Marthaler



Mathematics++ is a concise introduction to six selected areas of 20th century mathematics providing numerous modern mathematical tools used in contemporary research in computer science, engineering, and other fields. The areas are: measure theory, high-dimensional geometry, Fourier analysis, representations of groups, multivariate polynomials, and topology. For each of the areas, the authors introduce basic notions, examples, and results. The presentation is clear and accessible, stressing intuitive understanding, and it includes carefully selected exercises as an integral part. Theory is complemented by applications—some quite surprising—in theoretical computer science and discrete mathematics. The chapters are independent of one another and can be studied in any order.

It is assumed that the reader has gone through the basic mathematics courses. Although the book was conceived while the authors were teaching Ph.D. students in theoretical computer science and discrete mathematics, it will be useful for a much wider audience, such as mathematicians specializing in other areas, mathematics students deciding what specialization to pursue, or experts in engineering or other fields.

ISBN: 978-1-4704-2261-5

9 781470 422615

STML/75

For additional information and updates on this book, visit
www.ams.org/bookpages/stml-75

AMS *on the* Web
www.ams.org